DOCUMENT RESUME

ED 423 253 TM 028 988

TITLE	Improving Classroom Assessment: A Toolkit for Professional Developers (Toolkit 98).
INSTITUTION	Northwest Regional Educational Lab., Portland, OR.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1998-02-00
NOTE	698p.
CONTRACT	RV96006501
AVAILABLE FROM	Northwest Regional Educational Laboratory, 101 SW Main Street, Suite 500, Portland, OR 97204; fax: 503-275-0458; http://www.nwrel.org. Available in both print and CD-ROM versions.
PUB TYPE	Guides - Non-Classroom (055) -- Non-Print Media (100)
EDRS PRICE	MF04/PC28 Plus Postage.
DESCRIPTORS	*Educational Assessment; Elementary Secondary Education; *Grading; Optical Data Disks; *Performance Based Assessment; *Standards; Student Evaluation; *Test Construction; Test Use
IDENTIFIERS	*Alternative Assessment

ABSTRACT

This Toolkit is a guide to assessing student achievement and making students partners in the educational enterprise by clearly defining expectations for performance and letting students in on them. It is essentially about performance-based instruction and the role of student assessment in guiding and invigorating practice. This volume, Volume I, contains text and professional development training activities organized into four chapters. The first chapter, "Standards-Based Assessment--Nurturing Learning," reviews current thinking about standards-based instruction and the role of ongoing assessment. Chapter 2, "Integrating Assessment with Instruction," helps the reader understand the various ways that the development and use of assessment can affect and enhance instruction. The emphasis is on performance assessment. Chapter 3, "Designing High Quality Assessments," provides a summary and analysis of current alternative assessment efforts and when to use various designs. In Chapter 4, "Grading and Reporting--A Closer Look," why, whether, and how to grade students is explored. Each chapter contains an introduction that defines its goals, an outline of content, and an index of the professional development activities included in the chapter. A written section presents information on the concepts and ideas in the chapter, and associated professional development activities focus on implementing these concepts. Volume Two contains five appendixes with the supplemental resource material needed for various activities. It is packaged separately because the same material may be used for several activities. A version of the Toolkit on CD-ROM is included with Volume I. Volume I contains numerous overheads. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Improving Classroom Assessment:

A Toolkit for Professional Developers

Toolkit98

ED 423 253

TM028988

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Jerry Kirkpatrick
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

February 1998

By: Regional Educational Laboratories

Sponsored by the U.S. Department of Education
Office of Educational Research and Improvement (OERI)

BEST COPY AVAILABLE

Improving Classroom Assessment: A Toolkit for Professional Developers

(*Toolkit98*)

February 1998

By:

Regional Educational Laboratories

Availability:

First check your Regional Educational Laboratory.

Also available centrally from:
Northwest Regional Educational Laboratory
101 SW Main Street, Suite 500
Portland, OR 97204
(503) 275-9519; Fax: (503) 275-0458; http://www.nwrel.org

Sponsored by the U.S. Department of Education
Office of Educational Research and Improvement (OERI)

Regional Educational Laboratories

Appalachia Educational Laboratory (AEL)
Charleston, West Virginia

The LAB at Brown (LAB)
Providence, Rhode Island

Laboratory for Student Success (LSS)
Philadelphia, Pennsylvania

Mid-continent Regional Educational Laboratory (McREL)
Aurora, Colorado

North Central Regional Educational Laboratory (NCREL)
Oak Brook, Illinois

Northwest Regional Educational Laboratory (NWREL)
Portland, Oregon

Pacific Resources for Education and Learning (PREL)
Honolulu, Hawaii

SouthEastern Regional Vision for Education (SERVE)
Greensboro, North Carolina

Southwest Educational Development Laboratory (SEDL)
Austin, Texas

WestEd
San Francisco, California

This publication is sponsored wholly or in part by the U.S. Department of Education, Office of Educational Research and Improvement, under contract number RJ96006501. The content of this publication does not necessarily reflect the views of any individual laboratory, the Department of Education, or any other agency of the U.S. Government.

© Copyright, Northwest Regional Educational Laboratory, 1998, all rights reserved. Permission is granted to other RELs for reproduction and sale. Permission is granted to educators to copy portions for use in local professional development activities.

Acknowledgments

Many individuals outside the Regional Educational Laboratories contributed time and ideas during production of *Toolkit98* and its previous editions. We would especially like to acknowledge the following:

Thanks to Jay Millman (Cornell University), Jay McTighe (Maryland Assessment Consortium), and Elliott Asp (Littleton School District, Colorado) for reviewing a 1993 draft of the Toolkit and providing ideas for future directions.

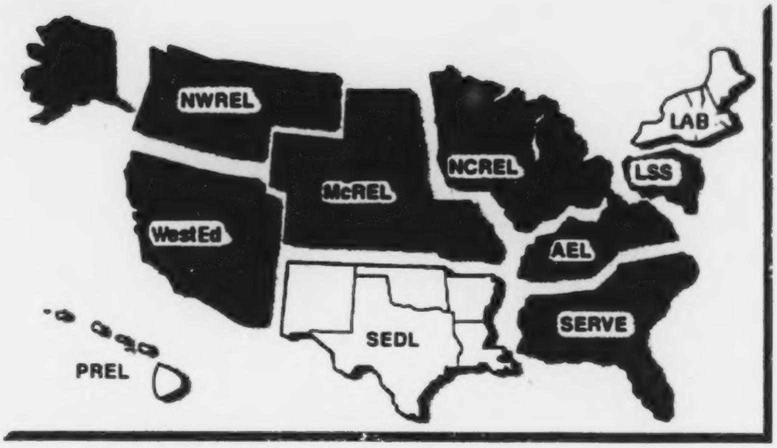
Thanks to Joe Hansen (Colorado Springs Public Schools, Colorado), Jan Chappuis (Central Kitsap School District, Washington), Steve Nolte (Ft. Hayes Educational Development Center, Kansas), Lorna Earl (University of Toronto, Ontario, Canada), Elliott Asp (Douglas County Schools, Colorado), and Mary Kay Armour (Maryland Assessment Consortium) for reviewing *Toolkit98* and providing many useful suggestions for making it better.

Thanks to all those authors and organizations willing to share their efforts by allowing us to reprint portions of their alternative assessments for **Appendix A**: Bellingham Public Schools (Washington), Wauwatosa School District (Wisconsin), Rebus Inc. (Ann Arbor, Michigan), Juneau Borough School District (Alaska), Multnomah Educational Service District (Portland, Oregon), Utah Department of Education, Illinois State Board of Education, Maryland Department of Education, National Assessment of Educational Progress (Princeton, New Jersey), Kentucky Department of Education, Mathematical Sciences Education Board (Washington, DC), New Standards Project (Pittsburgh, Pennsylvania), Lake County Educational Service Center (Grayslake, Illinois), Kansas State Department of Education, Oregon Department of Education, Massachusetts Department of Education, Martha Kinney (Journal of Reading), University of California at Santa Barbara, John Zola (Fairview High School, Boulder, Colorado), Alberta Education (Canada), Wisconsin Center for Educational Research (Madison, Wisconsin), Lake Oswego School District (Oregon), Curriculum Research and Development Group (CRDG), University of Hawaii (Honolulu), Social Science Education Consortium (Boulder, Colorado), State University of New York (Buffalo), Holt Public Schools (Michigan), California Department of Education, National Center for Research on Evaluation, Standards, and Student Testing (UCLA, Los Angeles, California), Bank St. College of Education (New York), R.A. Long High School (Longview, Washington).

Thanks to those authors allowing us to reprint their articles for **Appendix C**: Alfie Kohn, the Griffin Center, Pamela Brown Clarridge, ASCD, Douglas J. MacIver, Russell Gersten, Robert Lynn Canady, Tina Juarez, Lynn Olson, Heidi Watts, David Frisbie, and Grant Wiggins.

Finally, we greatly appreciate those individuals from organizations outside the RELs that contributed parts of activities/refinements on activities: Jan Chappuis, Activity 1.2 (Central Kitsap District, Washington); Joan Baron, Activity 1.4 (Harvard University); Ruth Sutton,

Activity 1.5 (Salford, England); Maryland Assessment Consortium, Activity 1.5 (Ijamsville); Joe Hansen, Activity 1.9 (Colorado Springs Public Schools); David Clarke, Activity 3.4 (University of Melbourne, Australia); Linda Elman, Activity 4.6 (Central Kitsap School District, Washington); Ken O'Connor, Activity 4.8 (Scarborough Board of Education, Ontario, Canada); Susan Brookhart, Activity 4.9 (Duquesne University).



Network Participants

Sharon Horn
OERI
555 New Jersey Ave., NW
Washington, DC 20208
(202) 219-2203
Sharon_Horn@ed.gov

Jane Hange
AEL
PO Box 1348
Charleston, WV 25301
(304) 347-0476
hangej@ael.org

Joan Herman
CRESST/CSE UCLA
301 GSE&IS/Mailbox 951522
Los Angeles, CA 90095
(310) 206-1532
joan@cse.ucla.edu

Jeong-Ran Kim
LSS
Temple University
9th Floor Ritter Annex
13th & Cecil B. Moore Ave.
Philadelphia PA 19122
(215) 204-2105
jkim3@vm.temple.edu

Eileen Ferrance
LAB
222 Richmond St., Suite 300
Providence, RI 02903
(401) 274-9548
eileen_ferrance@brown.edu

Don Burger
McREL
2500 Parker Rd., Suite 500
Aurora, CO 80014
(303) 337-0991
dburger@mcrel.org

Cathy Cook
NCREL
1900 Spring Rd., Suite 300
Oak Brook, IL 60521
(630) 571-4700
cook@ncrel.org

Judy Arter
NWREL
101 SW Main, Suite 500
Portland, OR 97204
(503) 275-9562
arterj@nwrel.org

Kathy Busick
PREL
828 Fort St. Mall, Suite 500
Honolulu, HI 96813
(808) 533-6000
busickk@prel.hawaii.edu

Chris Ferguson
SEDL
211 E. 7th St.
Austin, TX 78701
(512) 476-6861
cferguso@sedl.org

Nancy McMunn
SERVE
PO Box 5367
Greensboro, NC 27435
(336) 334-3211
nancy.mcmunn@bbs.serve.org

Mahna Schwager
Stanley Rabinowitz
WestEd
730 Harrison St.
San Francisco, CA
(415) 565-3000
mschwag@wested.org

BEST COPY AVAILABLE

Authors/Lab Contributors

Lab	Individuals	Authored or Contributed to:
AEL	Carolyn Carter, Jane Hange	Chapter 3 text; Activities 1.1, 1.4; Sample A.9
CRESST*	Joan Herman	Chapters 1, 2, and 3 text; Appendix E
LAB	Eileen Ferrance	Cover design
LSS	David Bechtel	Activity 1.8
McREL	Don Burger, Alice Krueger	Introduction; Activity Intro.1
NCREL	Cathy Cook, Linda Bond	Cover design; Chapter 1 text; Activity 1.8; Appendix E
	Melissa Chapko	Cover design
NWREL	Judy Arter	Introduction; Activity Intro.1; Chapter 1 text; Activities 1.2, 1.3, 1.5, 1.6, 1.7, 1.9, 1.10; Chapter 2 text; Activities 2.1, 2.5; Chapter 3 text; Activities 3.1, 3.3, 3.9; Chapter 4 text; Activities 4.2, 4.3, 4.6, 4.9; Appendix A Indexes; Samples A.1-A.8, A.10-A.24, A.26-A.30, A.32-A.48; Appendix B Index; Samples B.1, B.3, B.4, B.6; Appendix C; Appendix D agendas; Appendix E
	Kathy Petersen	Production
PREL	Kathy Busick	Activities 1.1, 1.3, 1.4, 1.5, 1.12, 2.2, 3.4, 3.7, 3.8, 4.2, 4.4, 4.7; Samples A.31 and B.5; Appendix D agendas
SEDL	Chris Ferguson, Glenda Clark, Jackie Palmer, Barbara Salyer,	Activity Intro.1; Chapter 1 text; Activities 1.1, 1.9, 2.4; Chapter 3 text; Activities 4.4, 4.5; Sample B.2; Appendix D agendas
SERVE	Nancy McMunn, Wendy McColskey	Introduction; Activities Intro.2, 1.5, 2.3, 3.2, 3.6, 4.1
WestEd	Mahna Schwager, Ann Muench, Kirsten Daehler	Chapter 1 text; Activities 1.4, 3.5; Samples A.25, B.7; general editing
	Joy Zimmerman	General editing

*National Center for Research on Evaluation, Standards, and Student Testing, a national research center funded, in part, by the federal government to conduct research and development on student assessment.

Toolkit98 Table of Contents

This Table of Contents probably isn't what you expect—it's not a nice list of all major sections with page numbers. Since each section of *Toolkit98* is numbered separately (so that it is easy to add, delete, and move material around without having to renumber the whole thing), this Table of Contents merely provides brief descriptions of chapters, activities, and appendices. The **Introduction** contains information on how to find what you want in *Toolkit98*. There is also a table of contents at the beginning of each chapter.

Introduction: Navigating *Toolkit98*

This chapter covers how to find what you want in *Toolkit98*—tabs, footers, chapter content and numbering, appendices, etc. It also includes *Toolkit98* overview information—target audience, goals, etc.

Activity Intro.1 *Toolkit98 Scavenger Hunt*

Helps users to learn how to find what they want in *Toolkit98*. Time: 30-45 minutes

Activity Intro.2 *Creating an Assessment Vision: Building Our Barn*!

Assists participants to understand how assessment change occurs by having them view the necessary components of change. Time: 1½ to 2 hours

Chapter 1: Standards-Based Assessment—Nurturing Learning

This is the "big picture" chapter—why careful attention to student assessment is crucial and how assessment activities fit into and support other current changes in education. It can also be thought of as the "awareness" chapter—introducing basic concepts, exploring current attitudes, and building a vision of what we want assessment to do for us.

Activity 1.1 *Changing Assessment Practices—What Difference Does it Make for Students?*

This activity stimulates thinking and discussion among teachers, parents, and diverse communities of learners about why assessment practices are changing. It is a good introductory activity. Time: 20-30 minutes

Activity 1.2 *Clear Targets—What Types Are These?*

In this activity, participants consider different types of learning targets for students, having a good mix of targets, and ensuring that learning targets are crystal clear. It is an intermediate level activity. Time: 40-60 minutes

Activity 1.3 *Post-it Notes™*

This activity reviews assessment terminology and serves as a mixer. It uses the glossary in **Appendix E**, and is a good introductory activity. Time: 20 minutes

Activity 1.4 *Seeing Wholes*

This activity emphasizes the connection between assessment, curriculum, and instruction. It is a good introductory activity. Time: 20 minutes

Activity 1.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore both the meaning of quality with respect to alternative assessment and why we all should care about quality. It is designed for educators at an intermediate level of study about assessment. Time: 75 minutes

Activity 1.6 *A Comparison of Multiple-Choice and Alternative Assessment*

Participants compare a multiple-choice test to an alternative assessment that attempts to measure the same skills, and discuss the advantages and disadvantages of each approach. This is an introductory level activity which uses assessment samples in **Appendix A**. Time: 30-60 minutes

Activity 1.7 *Target-Method Match*

This activity introduces assessment methods, and gives participants practice in matching methods to learning targets. It is designed for educators at an intermediate level of study about assessment. Time: 60-90 minutes

Activity 1.8 *Sam's Story: Comprehensive Assessment*

This activity illustrates the need for multiple measures of student achievement. It is an introductory level activity. Time: 45 minutes

Activity 1.9 *Going to School*

Part A of this activity demonstrates the importance of performance criteria; Part B illustrates different types of performance criteria, and Part C discusses the advantages and disadvantages of various types of rubrics. The activity uses assessment samples in **Appendix A**. Part A is an intermediate activity; Parts B & C are advanced. Time: 60 to 90 minutes

Activity 1.10 *Clear Targets and Appropriate methods—The View From the Classroom*

Participants self-evaluate the extent to which their instruction is focused on clear targets and the degree of conscious matching they perform between targets and assessment methods. Self-rating rubrics are provided. This activity is intermediate in difficulty. Time: 30-45 minutes

Activity 1.11 *Assessment Standards*

This activity provides an opportunity for participants to look at assessment standards in various content areas and related examples of assessment questions. It is an intermediate level activity. Time: 90 minutes

Activity 1.12 *Assessment Principles*

Participants examine the beliefs that influence their decisions about student assessment and explore equity issues in assessment. It is an introductory activity. Time: 20 minutes

Chapter 2: Integrating Assessment With Instruction

This chapter considers the ways assessment influences teachers, instruction, and students. Such consideration is essential if we want to build an assessment system based on a clear vision of what we want assessment to accomplish.

Activity 2.1 *Sorting Student Work*

This activity is at an intermediate level of difficulty and involves sorting samples of student work into three stacks representing "strong," "medium," and "weak" responses to a performance task. Participants have the opportunity to describe and discuss the characteristics of work that differentiate these stacks. This is an exercise in developing performance criteria, but it also demonstrates how developing performance criteria can help sharpen understanding of the goals held for students, increase teacher expertise, and assist students with understanding how to produce high quality work. The activity includes a description of the steps for developing performance criteria, using self-reflection letters as a running example. The activity uses samples of student work in **Appendix B**, and sample assessments from **Appendix A.** Time: 75-120 minutes

Activity 2.2 *How Can We Know They're Learning?*

This beginning level activity is an adaptation for parents of **Activity 2.1**. Parents have the opportunity to sort student work and distinguish for themselves the characteristics that make it more or less strong. The activity also helps parents understand changes in assessment and why they are occurring. It uses sample assessments from **Appendix A**. Time: Part A, 75-90 minutes; Part B, 30-40 minutes

Activity 2.3 *Ms. Toliver's Mathematics Class*

In this activity, using a video, a teacher demonstrates how she conducts continuous monitoring of student achievement and implements standards-based assessment. It is designed for educators at an intermediate level in their study of assessment. Time: 1½ to 2½ hours

Activity 2.4 *Is Less More?*

Participants experience an integrated, interactive, standards-based mathematics and science lesson and consider the implications of such a lesson for student assessment. It is designed for educators at an intermediate level in their study of assessment. Time: 2 hours

Activity 2.5 *How Knowledge of Performance Criteria Affect Performance*

This activity illustrates the importance of performance criteria in helping students understand the requirements of an assignment or task, and illustrates what can happen when criteria for judging success are not clearly understood by students. It is designed for educators at an intermediate level in their study of assessment. Time: 30-45 minutes

Chapter 3: Designing High Quality Assessments

While **Chapters 1** and **2** strive to develop overall visions of the role of assessment in instruction and the implications of purpose for how assessments are designed, **Chapter 3** begins the detailed examination of design options and related quality considerations. What do some of the current assessments look like? When should we use various design options? How does assessment purpose relate to design options? What do high quality alternative assessments look like?

Activity 3.1 *Performance Tasks—Keys to Success*

This advanced level activity illustrates the dimensions along which performance tasks differ and discusses the advantages and disadvantages of different approaches. It uses example assessments from **Appendix A**. Time: 60 to 90 minutes

Activity 3.2 *Spectrum of Assessment Activity*

Part A of this activity asks participants to review Bloom's Taxonomy and develop assessment questions and tasks that tap different levels. In Part B, participants review a variety of short, related assessment tasks and rank them according to the type of skills and knowledge each might elicit from students. The goal is to design tasks that really assess desired learning targets. This is an advanced level activity. Time: Part A, 90 minutes to 2 hours; Part B, 90 minutes

Activity 3.3 *Performance Criteria—Keys to Success*

This advanced level activity illustrates the characteristics of sound performance criteria. It uses assessment samples from **Appendix A**. Time: 60-90 minutes

Activity 3.4 *Assessing Learning: The Student's Toolbox*

This intermediate activity illustrates the relationship between different ways to design assessment tasks and the student learning to be assessed. Time: 20-30 minutes (variation takes 45-60 minutes)

Activity 3.5 *Performance Tasks and Criteria: A Mini-Development Activity*

In this advanced level activity, participants discuss how good assessment equals good instruction and the issues in using and designing performance assessments. Participants use assessment samples from **Appendix A**. Time: 2-3 hours

Activity 3.6 *How to Critique an Assessment*

In this advanced level activity, participants practice aligning assessment and instruction and evaluating a performance assessment for quality. Time: 1-2 hours

Activity 3.7 *Chickens and Pigs: Language and Assessment*

This introductory level activity emphasizes the critical role that language plays in effective and equitable assessments. Time: 10-20 minutes

Activity 3.8 *Questions About Culture and Assessment*

This introductory level activity increases awareness of the relationship between culture and assessment by demonstrating how cultural factors can affect student ability to show what they know and can do. Time: 90 minutes

Activity 3.9 *Tagalog Math Problem*

This introductory level activity illustrates how lack of knowledge of a language can affect performance on skills that have nothing to do with language understanding (like math). Time: 20-30 minutes

Chapter 4: Grading and Reporting—A Closer Look

Teachers at all grade levels are currently feeling a certain amount of discomfort about the manner in which (either by choice or by district policy) they have assigned grades. There is a sense that something has to change, but no-one is really sure what needs to change to or even how to productively frame the questions to be addressed. No one currently has *the* answer. However, in this chapter, we provide ideas—current best thinking about the issues that have to be addressed and how others have addressed these issues in redesigning grading and reporting systems.

Activity 4.1 *Weighty Questions*

This advanced level activity helps to illustrate the importance of developing sound grading practices that reflect valued student learning targets. Time: 90 minutes to 2 hours; extension, 60 minutes

Activity 4.2 *Putting Grading and Reporting Questions in Perspective*

This intermediate level activity provides an opportunity for teachers to express their current questions and concerns about grading and relate them to three levels of concerns about grading proposed in a paper by Alfie Kohn in **Appendix C**. Time: 75 to 90 minutes

Activity 4.3 *Grading Jigsaw*

This intermediate level activity raises and promotes discussion about issues surrounding grading. It uses papers on grading and reporting in **Appendix C**. Time: 1 ½ to 2 ¼ hours

Activity 4.4 *Won't Some Things Ever Change?*

In this intermediate level activity, participants compare and critically examine the messages sent by report cards from 1916, 1943, 1965, and 1994. Time: 30-40 minutes.

Activity 4.5 *Report Card Formats*

This advanced level activity provides examples of innovative report card formats and promotes discussion of their relative advantages and disadvantages. Time 40-75 minutes

Activity 4.6 *How To Convert Rubric Scores to Grades*

In this advanced level activity, participants discuss the advantages and disadvantages of four procedures for converting rubric scores to grades. This case study focuses on writing. Time 60-75 minutes

Activity 4.7 *Case of the Killer Question*

This advanced level activity presents the real-life grading dilemmas faced by an alternative high school. The killer question is: "How do we capture and report the learning of our students without compromising the integration, authenticity, student autonomy and self-directed learning that's at the heart of our program?" Time: 1 ½ to 2 ¼ hours

Activity 4.8 *Guidelines for Grading*

This advanced level activity presents eight guidelines for grading taken from current research. Participants consider each guideline in light of the goals they have for grading. Time 1 ½ to 2 hours.

Activity 4.9 *Grading Scenarios*

This advanced level activity presents real-life grading dilemmas for discussion. Time: 90 minutes

Appendix A: Alternative Assessment Sampler

Appendix A contains material from 48 different assessment projects. Samples cover all grade levels and several content areas—reading, writing, social studies, mathematics and science. All samples have copyright clearance to use in training.

Appendix B: Student Work Samples

Appendix B has samples of student responses to various performance assessment tasks. A variety of grade levels and subject areas is represented. All samples have copyright clearance to use in training.

Appendix C: Articles

Appendix C contains several papers and articles about grading and reporting. All papers have copyright clearance to use in training.

Appendix D: Training Agenda Examples and Evaluation Forms

Appendix D provides sample training agendas and vignettes showing how others have used *Toolkit98* activities. **Appendix D** also includes sample *Toolkit98* and training evaluation forms.

Appendix E: *Glossary*

Appendix E provides a list of assessment-related words and what we've agreed they mean.

Introduction:
Navigating *Toolkit98*

Tookit98 is about assessing student achievement. But, for the consortium of authors who developed *Toolkit98*, it's really about much more than that. It's about the vision of success we have for *all* students and how we'll know when we get there. It's about making students partners in the educational enterprise by clearly defining our expectations for performance and letting students in on them. It's about encouraging student self-reflection and self-assessment. It's about developing assessments that demand real intellectual quality while honoring student diversity. And, it's about assisting educators to create a view of assessment that supports instructors and classroom instruction. In short, it's about performance-based instruction and the central role of on-going student assessment to guide and invigorate practice.

All of this is easy to say, and it sounds good, but what does it really mean in practice? Well, we can't promise that we have all the answers—but, we do have a certain amount of knowledge and experience to share that illustrates the integration of assessment and instruction and the role of assessment in enhancing student achievement.

Who It's For

Toolkit98 is designed to assist classroom teachers to become better assessors of student learning. The primary users of *Toolkit98* are intended to be those who have responsibility to coordinate and facilitate professional development in assessment for teachers.

Toolkit98 Goals

Specifically, the goals of *Toolkit98* are to:

1. Provide practical support for quality assessment.

2. Provide background/foundation information on alternative assessment.

3. Provide information on the types of assessments that are being developed around the country and world and illustrate these variations with actual samples.
4. Discuss the issues and considerations surrounding the development and use of alternative assessments so that educators become good consumers and users of current instruments and good developers of new assessments.
5. Emphasize the instructional potential as well as the monitoring functions of alternative assessment.
6. Assist users to develop a vision of what they want their assessments to look like and have reasons for these choices.
7. Provide professional development activities that allow educators to construct their own understandings of the nature and role of assessment in promoting, enhancing, and measuring student learning and achievement.

Although our emphasis is on assessment that takes place in the classroom, many of these topics are appropriate both for classroom assessment and for assessment that occurs across classrooms and throughout districts and states (large-scale assessment). We will note when a particular consideration or practice is more appropriate for large-scale or classroom assessment.

Definition

In *Toolkit98*, we use the following definition of *alternative assessment:*[*] *Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.*

Some people refer to this type of assessment as "constructed response."

Even though in this document we emphasize alternative assessment, we do not want to imply that only alternative assessments are worthwhile and all previous assessment efforts are worthless. Actually, no single assessment approach can accomplish all our goals. A balance must be built by carefully considering all the types of things to be assessed and the reasons to assess them, and then choosing the assessment approach that best matches these targets and purposes. Sometimes the answer will be an alternative assessment, sometimes traditional assessment, and sometimes a combination of the two.

---

[*] An interesting historical note is that 60 years ago multiple-choice was considered an alternative to the current assessment of the time—subjective, teacher-based judgment. The attempt was to be more "objective" by reducing subjective judgment. Now, "alternative" is used to mean assessments other than multiple-choice, true/false, and matching. However, the issue of being "objective" is still the same as 60 years ago. The goal now is to make subjective assessment (performance and other alternative assessment) as objective as possible.

Toolkit98 Content

Toolkit98 includes two volumes. **Volume One** contains text and professional development training activities organized into four chapters:

Chapter 1: Standards-Based Assessment—Nurturing Learning. In this chapter we begin with current thinking about standards-based instruction and the role of ongoing assessment of student skills and knowledge to inform practice. Then we take a quick survey of what it means to assess well—high quality for all kinds of assessment from the classroom to the boardroom; from multiple-choice to portfolios. Along the way we visit the reasons that changes in assessment are taking place, discuss purposes of assessment, look at the need for clear student learning targets, and provide help with deciding when alternative assessment should be used.

Chapter 2: Integrating Assessment With Instruction. The goal of this chapter is to assist the reader in understanding the various ways that development and use of assessment can affect and enhance instruction. The chapter includes building the vision of how performance assessments can be useful instructional tools if they are designed properly.

⊂ Related *Toolkit98* Chapters and Activities:

Activity Intro.1—*Scavenger Hunt* helps users become familiar with *Toolkit98* format and content.

Chapter 3: Designing High Quality Assessments. What's out there and how good is it? This chapter provides a summary and analysis of current alternative assessment efforts and when to use various designs. No discussion of options would be complete without a look at quality—what do good assessments look like? So this chapter builds on the notions of quality from **Chapter 1**—this time focusing just on alternative assessments. The guidelines for high quality alternative assessments presented in the chapter can be used when developing *or* selecting assessments. Samples from real assessment instruments illustrate the points made.

Chapter 4: Grading and Reporting—A Closer Look. Grading and reporting are two issues that consistently recur in discussions of classroom assessment. This chapter discusses the issues of why, whether, and how we should grade students, and other ways to report student progress besides grades.

BEST COPY AVAILABLE

Each chapter has several parts:

- A chapter introduction that includes goals for the chapter, an outline of chapter content, and an index of the professional development activities included in the chapter.
- A written section ("Readings") that presents information on the concepts and ideas in the chapter. These can be used as background reading for *you* (the professional developer), or as handouts for training participants.
- Associated professional development activities complete with presenter's outline, handouts, and hard copies of overheads.

Volume Two contains supplemental resource material needed for various training activities. It is packaged separately because the same material may be used for several different activities, and past toolkit users tell us it's easier to have it separate.

Appendix A—*Alternative Assessment Sampler.* **Appendix A** contains material from 48 different assessment projects. Samples cover all grade levels and several content areas—reading, writing, social studies, mathematics and science. All samples have copyright clearance to use in training.

Appendix B—*Student Work Samples.* **Appendix B** has samples of student responses to various performance assessment tasks. Several grade levels and subject areas are represented. All samples have copyright clearance to use in training.

Appendix C—*Articles.* **Appendix C** contains papers and articles about grading and reporting. All papers have copyright clearance to use in training.

Appendix D—*Training Agenda Examples and Evaluation Forms.* **Appendix D** includes sample training agendas that illustrate how various activities in *Toolkit98 could* be sequenced, along with sample *Toolkit98* and training evaluation forms.

Appendix E—*Glossary.* This provides a list of assessment-related words and what they mean.

Use of *Toolkit98*

As the name implies, a toolkit contains a number of different tools that can be used to construct a multitude of products. The materials contained in *Toolkit98: Alternative Assessment* are like the hammers, saws, and screwdrivers of a carpenter's toolkit. They are mix and match; the user must choose the parts that will accomplish what is needed. When used creatively and with other necessary elements, the tools can build a useful and successful professional development experience. *Toolkit98* is not, however, intended to be a complete textbook on assessment. Nor is it intended to provide *all* the necessary background and expertise to accomplish effective professional development in assessment; it is intended to complement, not substitute for, local professional development and assessment expertise. Although we try to present highly technical topics in simple and accessible ways, facilitators and trainers will need technical expertise and experience to address the multitude of questions that often arise during professional development activities.

📖 **Reference Box–Related Products**

The *Promising Practices in Assessment Database* is available on the Internet at: http://assessment.wested.org/ppad/. This website is a searchable database of assessment resources available from all the Regional Educational Laboratories.

***Facilitating Systemic Change in Science and Mathematics: A Toolkit for Professional Developers* (1995) is a set of learning activities that helps those supporting reform efforts deepen their knowledge about science and mathematics education, dissemination, professional development, and the change process. This resource was written by the Regional Educational Laboratories. Call your Regional Lab for information.**

And, although the professional development activities included in *Toolkit98* are designed to model effective teaching strategies that incorporate what we know about sound instructional practices, they may need to be tailored to fit local situations. Purpose, setting, participants, time, and resource considerations will affect the actual planning and implementation of a professional development opportunity. So please feel free to modify exercises and activities to meet your own needs.

BEST COPY AVAILABLE

Of Footnotes, Headers, and Other Navigational Hints

There's a lot of material in *Toolkit98*. We've included some features to help users find what they want and then, when they're done, replace it where it goes. Here are some hints:

Tabs

In each chapter there is a tab that shows where the chapter itself begins and another one to show where the activities for that chapter begin. There are also tabs in the appendices at every fifth sample, so users don't have to dig through the whole stack to see where, for instance, sample 13 starts.

Footnotes and Page Numbering

Each chapter and activity is numbered separately. Thus, **Chapter 1** text begins on page 1, but so do **Activities 1.1, 1.2**, ...and so on. Reason—so we don't have to renumber everything anytime we want to add or change a single page. But, we have wonderful footers that will help you find what you want:

Toolkit98: Chapter X **Text**—*Title* page #

(*e.g., Toolkit98:* Chapter 3 **Text**—*Designing High Quality Assessments* 10)

Toolkit98: Chapter X **Activity X.Y**—*Title* page #

(e.g., *Toolkit98:* Chapter 4 **Activity 4.1**—*Weighty Questions* 1)

Toolkit98: Appendix α **Sample** α.**Y**—*Title* page #

(e.g., *Toolkit98:* Appendix B **Sample B.1**—*Primary Writing* 3)

Where: X = Chapter number (from 1 to 4)
Title = the title of the chapter, activity, or sample
α = appendix letter
Y = the particular activity or sample number within a chapter or appendix

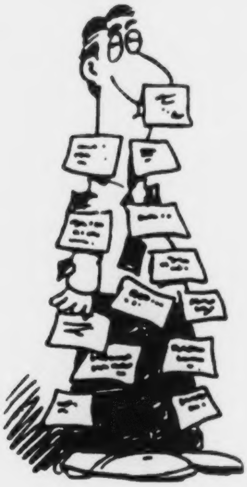
So, let's say that **Activity 4.1** is removed for use. When done, the footer tells the user that the material belongs in **Chapter 4** after the tab for "Activities"; and it's the first activity.

Icons and Sidebars

1. To help users find activities that are related to topics addressed in the text of each chapter, we've added sidebars and an icon, like that to the right.
2. Instead of having endless lists of references at the end of each chapter, we have been selective, and those we cite are included as a sidebar in the text—for example, the reference box on page 5.
3. We've used sidebars when we want to make a point related to the text, but it might interfere with the overall flow. A sidebar is used also when we want to make sure something is seen—for example the definition on page 2.

⊂ Related *Toolkit98* Chapters and Activities:

Activities Intro.1—*Toolkit98 Scavenger Hunt* **and Intro.2**—*Creating an Assessment Vision, Building Our Barn* **are included in this chapter.**



BEST COPY AVAILABLE

What We've Learned . . .

Toolkit98 is intended to assist users to design and develop professional development in assessment that engages participants in a practical hands-on fashion. We've learned a lot over the last 20 years about assessment training and promoting change. The list we humbly submit below probably matches up to your experience. We have attempted to design the *Toolkit98* text and professional development activities to reflect this knowledge and experience.

The major thing we've learned is this:

> **If we can't demonstrate how changes in assessment will make teaching and/or student achievement faster, easier, and better, then we should not embark on teacher assessment training. Corollary: If we believe that assessment has power in the classroom, we ought to be ready to demonstrate it.**

And now the rest of the list.

What we've learned . . about assessment and its role in educational change. . . .

- Assessment, curriculum, and instruction must all be aligned with standards to have maximum impact.
- Connectivity—everything (assessment, instruction, reform, content standards, the change process, etc.) is related to everything else, so sometimes it is difficult to know where best to start when we begin updating educational practice.
- The process of changing educational practice is complex (because of it's interconnectedness), so everyone must merely…
- ....Start somewhere—take "baby steps" toward change.
- Assessment is an excellent place to start because it tugs at issues that seem to be at the heart of current recommended changes in practice. Assessment is where the rubber meets the road, where we really define what we want students to know and be able to do regardless of what we say in our content and performance standards. Assessment is also a direct indicator of student success—and so can command resources.
- It's best to start at the basic level with respect to assessment; few educators have had the opportunity to learn about assessment in a manner that is useful day to day in the classroom.

⊂ Related *Toolkit98* Chapters and Activities:

Activity Intro.2—*Creating an Assessment Vision: Building Our Barn* assists users to place assessment in the context of systemic reform and change.

23

- Assessments need a purposeful design to promote student success—good assessment involves more than just giving a test or collecting data.
- Assessment is not a glamorous activity; it's hard, yet rewarding, work.
- Assessment terms are used differently in different places. There is no universal agreement on the meaning of terms, so discussions can get confusing. Therefore, no one should claim to use the "right" definitions while everyone else uses the "wrong" definitions.

What we've learned . . . about professional development in assessment. . .

- Professional development in assessment needs to be job-embedded and be undertaken with a long term commitment because assessment reform is like any other reform: all steps of the change process apply. Anyone who thinks that changing assessment will be a quick-fix for all educational woes is wrong. Therefore…
- ...One can't change beliefs and attitudes in a single workshop—teachers must go through the process. So…
- ...Professional development in assessment must be conceived of as more than a series of workshops (or worse yet, a single workshop). So…
- ...Training in assessment must be supported at all levels.
- Merely changing assessment will not improve student learning. Teachers need to experience how changes in assessment relate to modifications in day-to-day instruction. So…
- ...Assessments must be designed to impact classroom practices and then teachers must come to their own understandings of what practices to change and how to change them.

What we've learned. . . about how adults learn. . . .

- Teachers (and, indeed, all adults) learn the same way that students learn. They have a variety of learning styles, need to apply their knowledge to relevant real-life situations, need to practice with feedback, and proceed through a learning curve or developmental continuum.
- The learning process is reiterative—there is always more to learn.
- The learning process must promote understanding. Plans that call for providing teachers with canned materials will not promote meaningful change. Teachers need to know, understand, and embrace suggested changes, not blindly use things they don't understand or deem important.

What we are saying is that the steps involved in changing assessment practice are not new. Any type of change is hard, and planning is vital. Schools and districts that are successful over the long term are those that have taken the time to talk, plan, and consider many of the components listed above. In other words, they have a vision for success and a carefully thought out, purposeful plan for achieving it.

Activity Intro.1

Toolkit98 Scavenger Hunt

Purpose:

1. To help users explore the contents of *Toolkit98*
2. To help identify areas of participant interest or assessment needs that might be addressed in *Toolkit98*

Uses:

This activity introduces participants to the contents of *Toolkit98*. It can be used alone or in conjunction with an activity in which groups are assigned to prepare and present five-minute overviews of *Toolkit98* chapters. Both approaches are an alternative to the facilitator presenting chapter information. The activity can be done individually or in small groups.

Rationale:

Toolkit98 developers acknowledge that participants have different learning styles. Some, after getting *Toolkit98*, will immediately flip through the pages to see what's in it. Others will turn to the Table of Contents first. Still others may find it so overwhelming that they don't know where to begin. Yet most people who will use *Toolkit98* are pressed for time and appreciate getting a complete sense of the contents before leaving a workshop.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead AIntro.1,O1**—*Toolkit98 Scavenger Hunt Purposes*
- **Handouts AIntro.1,H1**—*Toolkit98 Scavenger Hunt;* **AIntro.1,H2**—*Answers to Toolkit98 Scavenger Hunt*

- Optional set of 5-6 small "prizes" (e.g., pencils, pens, candy, buttons) for participant groups that finish first and/or correctly

Time Required:

30-45 minutes

Facilitator's Notes:

1. Use **Overhead AIntro.1,O1** to present the purpose of the activity.

2. Allow 20 minutes for individuals or small groups to find the items listed in **Handout AIntro.1,H1**—*Toolkit98 Scavenger Hunt.*

3. As groups finish, you may pass out **Handout AIntro.1,H2**—*Answers to Toolkit98 Scavenger Hunt* for participants to check their work. (Note: This is an assessment that appropriately uses predominantly single "correct" answers.) You may wish to have a small "prize" for the individual or group that finishes first or has the correct answers.

4. Facilitate a discussion about items that were most difficult to find, clarify topics, or expand on areas of interest to the group. (Note: This is an opportunity for you to identify specific areas of assessment your group may need to discuss further.)

Optional Procedure

Time Required:

20 minutes

Facilitator's Notes:

If you are limited by time, divide participants into six groups and assign each group five items (e.g., group 1 would find items 1-5; group 2, items 6-10, etc.). Everyone can try for the Bonus item. Allow 5-10 minutes for groups to complete the task before reporting out their results. Then distribute **Handout AIntro.1,H2** to corroborate their findings.

Toolkit98
Scavenger Hunt

Purposes:

1. To explore the contents of *Toolkit98*
2. To help identify areas of participant interest or assessment needs that might be addressed in *Toolkit98*

Overhead AIntro.1,O1

Toolkit98 Scavenger Hunt

1. Where can you find abstracts of *Toolkit98* chapters?

2. Where are design options for performance criteria?

3. If you remove the overhead and handout masters for **Activity 2.1**—*Sorting Student Work*, what chapter, activity, and page numbers tell where they should be replaced?

4. Where does *Toolkit98* address different approaches for grading?

5. Where can you find samples of student work?

6. How many professional development activities are in *Toolkit98*?

7. Where does *Toolkit98* explain the differences between reliability and validity?

8. Where can you find articles about grading and reporting issues?

9. Which sample large-scale mathematics assessment lists performance criteria having four dimensions (called analytical trait criteria)?

10. Where is the picture of a slice of time in a spiral approach to "Continuous Monitoring of Student Learning?"

11. Where does it explain why this package is called a toolkit?

12. Where can you find alternative options for communicating about student learning?

13. Where is the operational definition of alternative assessment used by *Toolkit98* developers?

Handout AIntro.1,H1, p. 1

14. Which chapter discusses using assessment as a tool for instruction in the classroom?

15. Where can you find a form to provide feedback to the developers on use of *Toolkit98?*

16. Where can you find questions to consider in examining the consequences of an assessment?

17. Which activity raises issues related to chicken legs?

18. Where can you find vignettes of long-term professional development projects focusing on alternative assessment?

19. Where are the seven goals of *Toolkit98*?

20. Where is there a discussion of "Five Keys to Quality Assessment?"

21. Where is there a discussion of assessment and the change process?

22. Where can seven steps to assigning a grade be found?

23. Which sample assessment came out of the New Standards Project?

24. Where can you find a form for evaluating the quality of alternative assessments?

25. Where does *Toolkit98* address the difference between open-response and open-ended tasks?

26. Which activity uses a videotape of classroom mathematics instruction?

27. Where can you find a description of the changes that drive the need for new assessment approaches?

28. Where can you find definitions associated with alternative assessment?

29. Where can a sample large-scale high school science portfolio assessment be found?

30. Which samples have developmental continuums appropriate for primary students?

Bonus: Who meets the Blooms?

Answers to *Toolkit98* Scavenger Hunt

1. **Introduction**, pages 3-4; also first page of each chapter
2. **Chapter 3** text, pages 20-29
3. **Chapter 2**, **Activity 2.1**, pages 6-14
4. **Chapter 4** text, pages 7-13
5. **Appendix B**, Samples B.1-B.7; **Appendix A**, Samples A.3, A.8, A.15, A.39, A.42, A.43
6. 35 in **Chapters 1-4**, plus 2 in the **Introduction**
7. **Chapter 3** text, page 6; **Appendix E**—*Glossary*, pages 5 and 7
8. **Appendix C**

9 Sample A.18 in **Appendix A**—*Sampler*

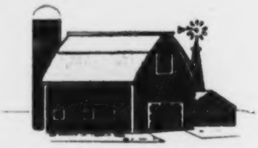
10. **Chapter 2** text, page 12
11. **Introduction**, page 5
12. **Chapter 4** text, pages 16-19; **Activity 4.5**—*Report Card Formats*
13. **Introduction**, page 2; **Chapter 3** text, page 4; **Appendix E**—*Glossary*, p. 2
14. **Chapter 2** text, pages 8-9 and 22-30
15. **Appendix D**, pages 16-17
16. **Chapter 3** text, page 36
17. **Activity 3.7**
18. **Appendix D**, pages 7-14
19. **Introduction**, pages 1-2
20. **Chapter 1** text, pages 14-30; **Activities 1.2, 1.7, 1.10**
21. **Activity AIntro.2**—*Building Our Barn*; **Chapter 1** text, pages 5-7; **Introduction**, pages 8-9
22. **Chapter 4** text, pages 6-16
23. **Sample A.15**—*Aquarium Problem* in **Appendix A**—*Sampler*
24. **Chapter 3 text, pages 40-41**; **Activity 3.6**, pages 11-12
25. **Chapter 3 text, pages 7-9**
26. **Activity 2.3**—*Miss Toliver*
27. **Chapter 1** text, pages 7-13
28. **Appendix E**
29. **Sample A.42**—*Science Portfolio* in ***Appendix A***—Sampler
30. **Appendix A**—*Sampler*, **Samples A.3, A.4, A.5**

Bonus: The Big Bad Wolf

Handout AIntro.1,H2

Activity Intro.2

Creating an Assessment Vision: Building Our Barn![1]



Purposes:

1. To discuss and recognize the components of a system
2. To explore the idea of systemic thinking
3. To reflect on systemic planning
4. To determine how to "build our barn"—building the "big picture" view for creating and implementing change at the classroom or school level

Uses:

This is an intermediate level activity best used after some assessment training has been done and there is a desire to begin to make more assessment changes at the classroom or school level. It may not be best to use this as an introductory activity, but as a reflective piece for further planning.

Rationale:

The best staff development opportunities for teachers are designed to be continuous, long-term and job-embedded. If teachers are to implement "quality" assessment practices they must focus on aligning curriculum, assessment, and instruction in order to get the "big picture" of what students know and are able to do. Teacher attitudes and beliefs about assessment do not change until teachers utilize the strategies of "quality" assessment with their students and actually see student learning improve. This activity helps educators create assessment improvement plans at the classroom, school or district level. It also explores ways to provide staff development that supports improvement goals for schools and professional development plans for teachers.

---

[1] Adapted from: Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers, available from the Regional Laboratory for Educational Improvement of the NE & Islands, 1995.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens, chart pack paper/note pads, marking pens
- **Overheads AIntro.2,O1**—*Creating a "Big Picture" for Change*, **AIntro.2,O2**—*Planning for Change*, **AIntro.2,O3**—*All of These Things Are Related...*, **AIntro.2,O4**—*A Collection of Related Parts...*, **AIntro.2,O5**—*Understanding Systems*, **AIntro.2,O6**—*How Did This Barn Get Built?*, **AIntro.2,O7**—*Looking at the Parts*, **AIntro.2,O8**—*Was the Barn Raising...?*, **AIntro.2,O9**—*Building Our Barn*, **AIntro.2,O10**—*Resources*, **AIntro.2,O11**—*Plans/Design and Instructions*, **AIntro.2,O12**—*Community Buy-In*, **AIntro.2,O13**—*Teamwork*, **AIntroO.2,O14**—*Leadership and Support*, **AIntro.2,O15**—*Trained People*, **AIntro.2,O16**—*A Shared Vision*, **AIntro.2,O17**—*Planning "Quality" Grade Level or School Assessment*, **AIntro.2,O18**—*Purposes*, **AIntro.2,O19**—*A Shared Vision*, **AIntro.2,O20**—*Resources*, **AIntro.2,O21**—*Student and Faculty Buy-In*, **AIntro.2,O22**—*Curriculum and Instruction*, **AIntro.2,O23**—*Leadership*, **AIntro.2,O24**—*Support*, **AIntro.2,O25**—*Trained People*, **AIntro.2,O26**—*Summary*
- **Handouts AIntro.2,H1**—*Building our Barn*, **AIntro.2,H2**—*School or Grade Level Assessment*, **AIntro.2,H3**—*Classroom Assessment*, **AIntro.2,H4**—*Components for Assessment Change*, **AIntro.2,H5**—*Assessment Action Planning Sheet, School or Grade*, **AIntro.2,H6**—*Assessment Action Planning Sheet, Classroom*
- **Video**—The movie *Witness* with Harrison Ford and Kelly McGillis

Reference Box

You may need to purchase an institutional copyright release to use the *Witness* video in training. Contact Paramount, 555 Melrose Ave., Hollywood, CA 90038.

Other useful background documents for looking at the change process in schools are:

Michael G. Fullan, *Why Teachers Must Become Change Agents*, *Educational Leadership*, March 1993, pp. 17.

Jerry L. Patterson, *Leading Through Systems Thinking*, in *Leadership for Tomorrow's Schools*, 1993, pp. 66-79.

Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers, Chapters 12 and 16, 1995, Regional Laboratory for Educational Improvement of the NE & Islands.

BEST COPY AVAILABLE

- Large display board—corkboard or wall space

Time Required:

1-1/2—2 hours for vision activity; planning afterwards can run 3+ hours

Facilitator's Notes:

A. (5-10 minutes) Set-up:

1. You need to have overheads in order, chart paper with markers, post-its on tables, a large display board, and these notes.

2. Introduce this activity by explaining the rationale. You might say:

   We are here today to plan for assessment training at your school—the components of the assessment process we want to consider to make this training fit into your school improvement plan and district goals.

   Use **Overheads AIntro.2,O1**—*Creating a "Big Picture" For Change,* and **AIntro.2,O2**—*Planning for Change*... to talk about change. Ask participants to look at the diagrams on **Overhead AIntro.2,O1** and discuss which arrow best represents their school's energy flow. Look at **Overhead AIntro.2,O2** which outlines the components needed to make change happen in a school, district, or classroom, and what happens when one component is missing or not supported. Ask participants to quickly audit their own situation with respect to each of these components.

B. (15-20 minutes) The Activity: Building Our Barn

1. **Define system**

   a. Using **Overhead AIntro.2,O3**—*All of These Things*...., ask:

      Do these related parts represent a system? (The answer is "no," because they only represent related parts and not a functioning whole.)

   b. Cover the lower words ("A working car is a system.") on the **Overhead AIntro.2,O4**—*A Collection of Related Parts*... and ask the participants if the car represents a working system. If so, why? Then, uncover the lower words and explain that the car is representative of a working system because the parts interact. ASK: *How might we define a system?*

   c. Use **Overhead AIntro.2, O5**—*Understanding Systems* to define "system."

d. Ask participants for examples of other systems they know, including systems in education. Ask them what the components are of some of the educational systems they know. (You may not get many answers; however, participants might cite curriculum, instruction, textbooks, assessment, administration, and so on.)

e. Using **Overhead AIntro.2,O6**—*How Did This Barn Get Built?* tell participants that they will begin to build their barns today. It could be a school, classroom or district barn they will begin to build, but in order to build a barn, they must understand how to look for all the components of a system. A 14-minute segment of the "Witness" video will be shown for this purpose.

2. (20 minutes) **Show *Witness* video**

a. Describe the segment they will see. *The "Witness" video segment is about a detective (played by actor Harrison Ford) who gets injured trying to protect a young Amish boy. The young boy had witnessed a murder committed by a high ranking police official. Ford's injury forces him to hide and recuperate at the Amish farm where the young boy and his mother (played by actress Kelly McGillis) live. This scene shows a recovered Ford, disguised as a member of the Amish village, helping to build a barn.*

b. Tell participants as they watch this video segment to write down (on paper) components of the barn building scene that reflect a working system. (**Overhead AIntro.2,O7**—*Looking at the Parts*). What is going on or had to have been in place in order to get this barn built?

c. Show the video segment. (The group will ask to watch the rest of the video. Resist the urge.)

3. (30 minutes) **Discuss the *Witness* Video**

a. Ask the group if this scene represented a working system (**AIntro.2,08**—*Was the barn raising...?*) (Yes, because "all the parts were functioning purposely for the benefit of the whole." Go back to the original definition of a system as needed.)

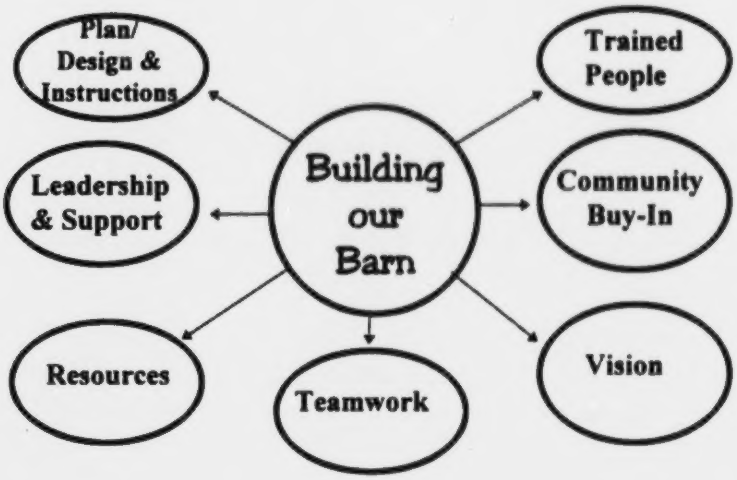
b. Ask participants to discuss in groups the components they wrote down.

c. In the large group, capture a component from *each* group and record it on chart paper. Return to each group and ask if they have another component to add that is not recorded. Continue to create the list until all the components are recorded.

d. Ask participants, in small groups, to try and group together related words that could be part of the same major component cluster. For example:

- "Purpose, a clear view of what you want and why you want it, a reason for building a barn and not, say, a garage or house, and what you want it to do for you," might be called **vision**
- "Hammer, nails, food, lumber, people, saws, etc.," might be called **resources**
- "Cooperation and people working together," might be called **teamwork**
- "Blueprint, model, plans, and instructions," might be called **plans**
- "Foreman and team heads," might be called **leadership**
- "Carpenters, plumbers, cooks, etc.," might be called **trained/skilled people**

e. Ask several groups to relate the major component clusters they created. Show participants that they basically came up with the same component clusters as other groups (**Overhead AIntro.2,O9**—*Building Our Barn*). Use **Overheads AIntro.2,O10** to **O16**, as needed to illustrate each bubble. **Handout AIntro.2,H1**—*Building Our Barn* also summarizes these component clusters.



4. (30 minutes) **Relate this to educational settings**

a. Ask participants to think about what they want to work on in assessment; for example, one group might want to work on assessment planning for their school, while another group might want to work on assessment planning at the district level, and another group might want to work on assessment planning for their classrooms. Break participants into these "like setting" groups.

b. Ask each group to discuss and come up with all component clusters in their setting that creates a systemic view of the "big picture." When they are done, ask them to compare their lists to the ones suggested in **Handouts AIntro.2, H2**—*School or Grade Level Assessment Changes* or **AIntro.2,H3**—*Classroom Assessment Changes*. Parallel overheads for **Handout AIntro.2, H2** are **AIntro.2,O17** to **O25**.

c. Compare the component clusters they created with the component clusters in the Barn Building activity and discuss the similarities and differences.

d. If their list of component clusters is different from those provided in **Handouts AIntro.2,H2** or **H3**, participants can create their own wheel of essential component clusters using **Handout AIntro.2,H4**—*Components for Assessment Change*.

5. (3+ hours) **Optional Planning Time**

a. Each group will choose one of the action planning sheets to complete in their groups (**Handout AIntro.2,H5**—*Assessment Planning - Creating a Vision for Your School* or **Handout AIntro.2,H6**— *Assessment Planning - Creating a Vision for My Classroom*). The School planning document requires a separate sheet for each component in the wheel. The facilitator should sit in on each group's discussions and offer ideas, as needed.

b. If time permits, groups should report on their progress. If not, suggest to each group that they complete the planning sheet back at their schools. At least this process will get them started thinking about the steps they must take if assessment changes are going to be a lasting part of their school or classroom environment. Suggest a follow-up session or time for you to visit the school to check on the progress of their planning.

6. (10 minutes) **Closure:**

Summarize things to remember when implementing assessment changes in a school or classroom (see **Overhead AIntro.2,O26**—*Summary*). Use these statements to reassure participants that assessment change is difficult and that when the changes involve the whole school it is vital to pay attention to the research on how change occurs.

- Planning and reflection time for teachers is very important.
- Assessment change is a slow process—take "baby steps." Focus on enhancing what is already being done. Do not try to make a lot of changes too fast.

- Change involves a sequence of concerns—awareness of the change, concern about the impact of the change on oneself, concern about implementing the change correctly, and, finally, the impact of the change on students. The stage of concern will determine what is done next and what support people need. (See the reference box for an Internet address that allows you determine the stages of concern about assessment for your staff.)

- Spend your time on the people who might lead the way for others—innovators and leaders. You might not be able to get some people to buy in to the idea of changing their assessment process. See the reference box for more information on personality types and their relationship to the change process.

- Promote an environment of "continuous growth through feedback spirals." (For more information about feedback spirals see the reference box.) Learning is never finished. If you want assessment change to become a lasting part of the district, school or classroom then it is vital to pay attention to:
    - Planning
    - Taking action and experimenting
    - Assessing and gathering evidence
    - Studying, reflecting, and evaluating
    - Modifying actions based on new knowledge
    - Revisiting goals and purposes

Reference Box

For descriptions of different personality types and their relationship to the change process see: Evelyn Rogers, *Diffusion of Innovations*, 1971, New York: The Free Press.

For more information on "continuous growth through feedback spirals," see Bena Kallick, *The Role of Assessment in the Learning Organization*, in ASCD Yearbook, *Communicating Student Learning*, Chapter 12, pp. 178-187, 1996, ASCD; or Bena Kallick and Art Costa (Eds.), *Assessment in the Learning Organization, Shifting the Paradigm*, 1995, Part II, pp. 25-28.

The Stages of Concern questionnaire is located on the Planet Innovation site and can be reached by first registering on the following page:

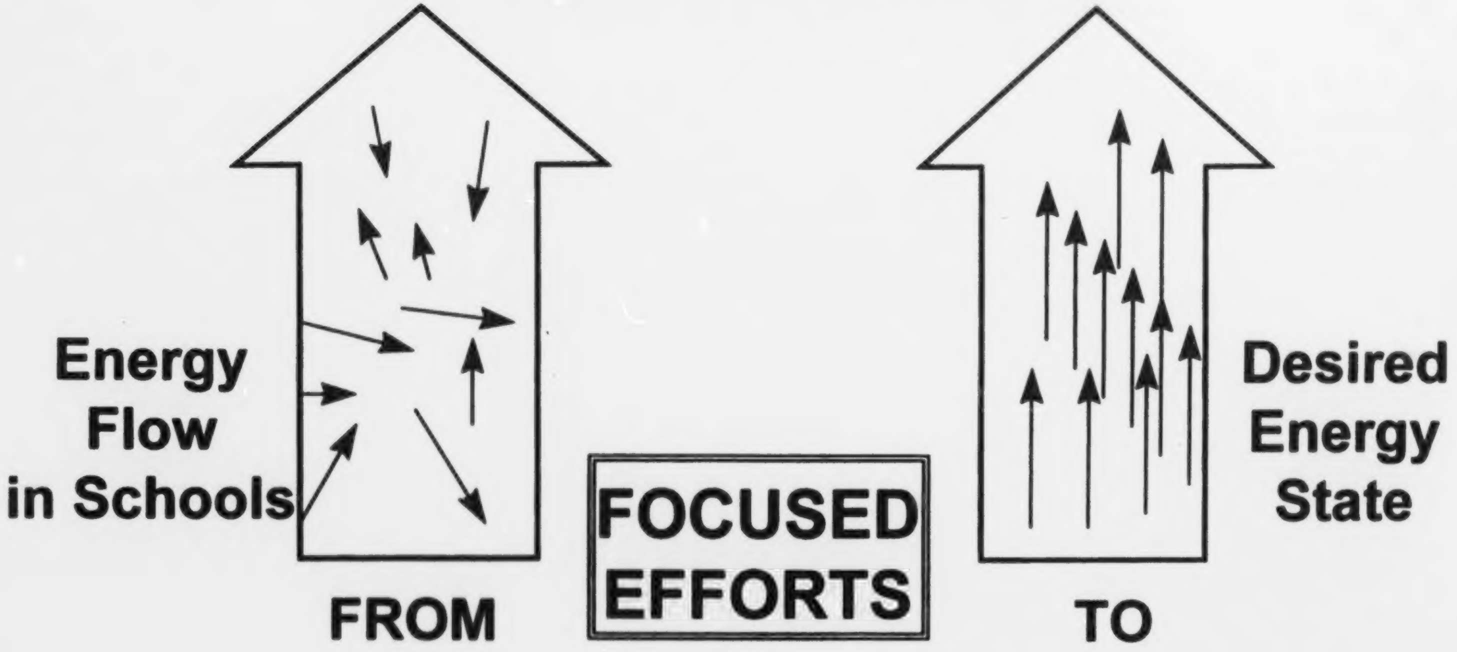
http://planet.rtec.org/

Once registered, you can go to the SoC page by following the links on the Planet Innovation site or directly at

http://planet.rtec.org/cbam/intro.cgi

BEST COPY AVAILABLE

CREATING A "BIG PICTURE" FOR CHANGE



Energy Flow in Schools

FROM

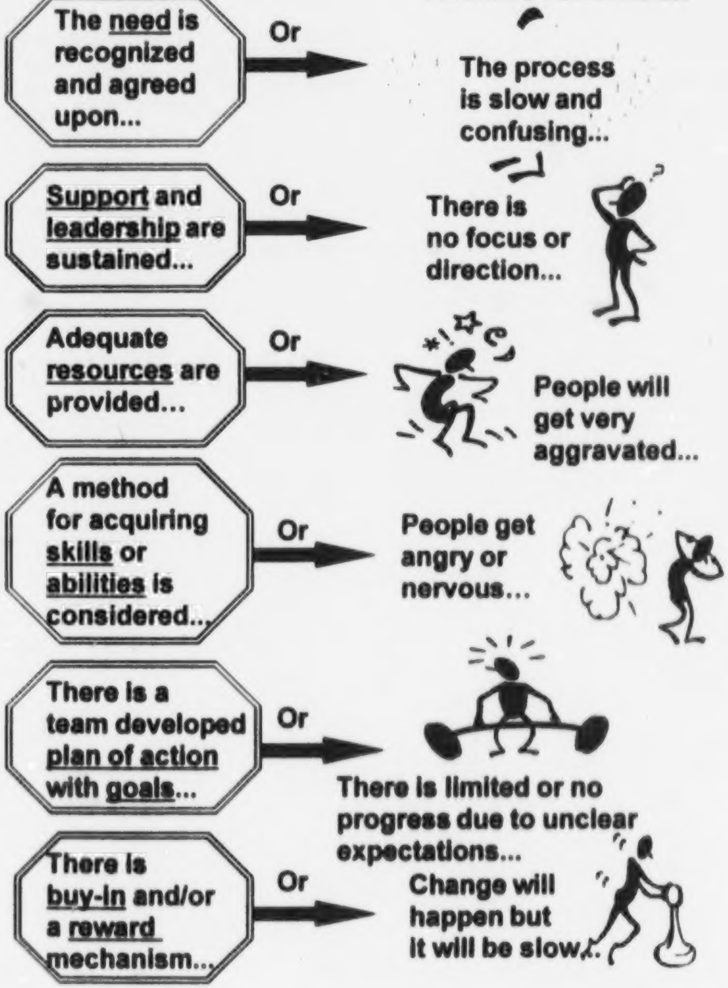
FOCUSED EFFORTS

TO

Desired Energy State

Overhead AIntro.2,O1

PLANNING FOR CHANGE



The need is recognized and agreed upon... Or → The process is slow and confusing...

Support and leadership are sustained... Or → There is no focus or direction...

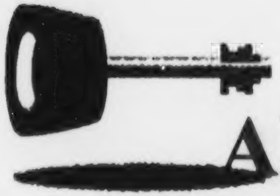
Adequate resources are provided... Or → People will get very aggravated...

A method for acquiring skills or abilities is considered... Or → People get angry or nervous...

There is a team developed plan of action with goals... Or → There is limited or no progress due to unclear expectations...

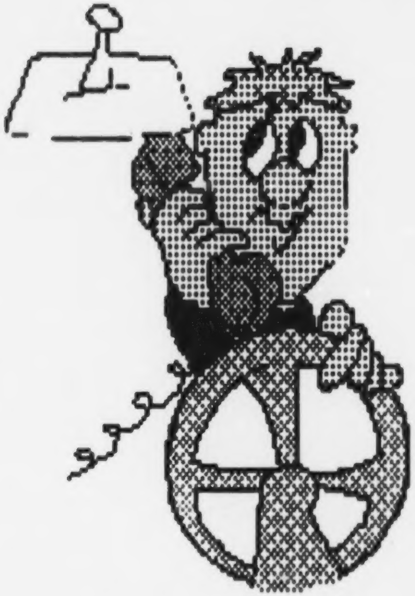
There is buy-in and/or a reward mechanism... Or → Change will happen but it will be slow...

Overhead AIntro.2,O2



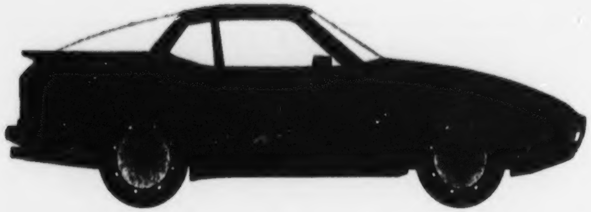
ALL OF THESE THINGS ARE RELATED TO A COMMON ITEM...

Do they represent a system?



Overhead AIntro.2,O3

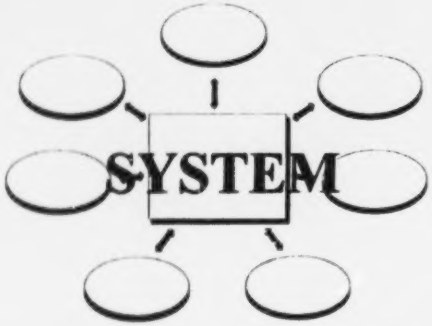
A collection of related parts is *NOT* a system.



A working car *IS* a system.

Overhead AIntro.2,O4

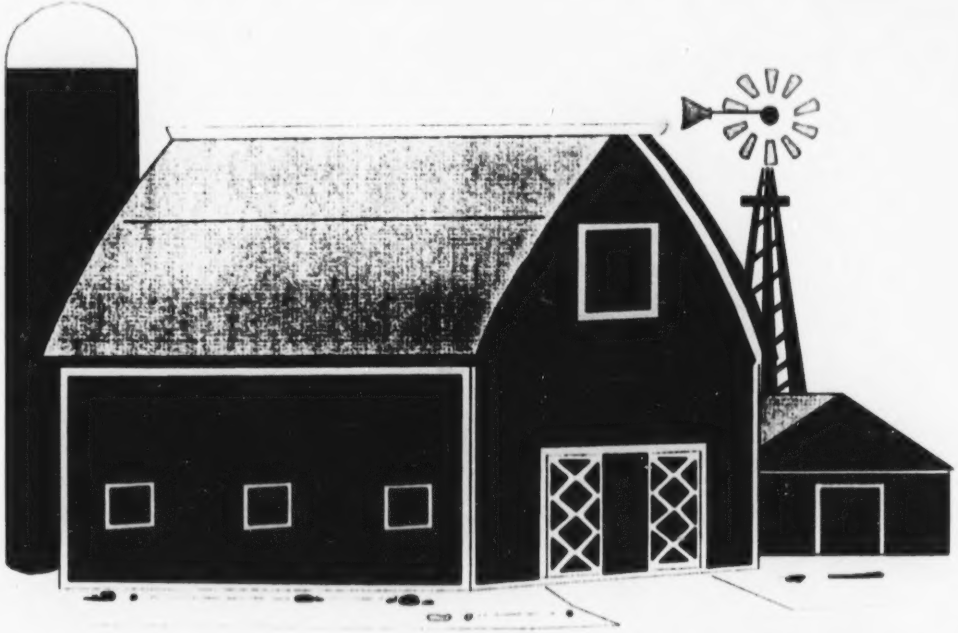
UNDERSTANDING SYSTEMS



A system is a collection of parts that interact to function purposefully as a whole.

Overhead AIntro.2,O5

HOW DID THIS BARN GET BUILT?



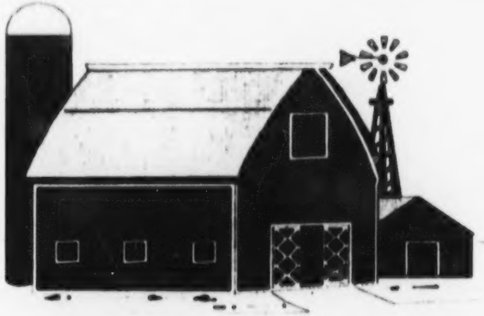
Overhead AIntro.2,O6

LOOKING AT THE PARTS.

- **While watching the video, “Witness,” think about the components (parts) of the BARN RAISING SCENE that reflect a WORKING SYSTEM.**

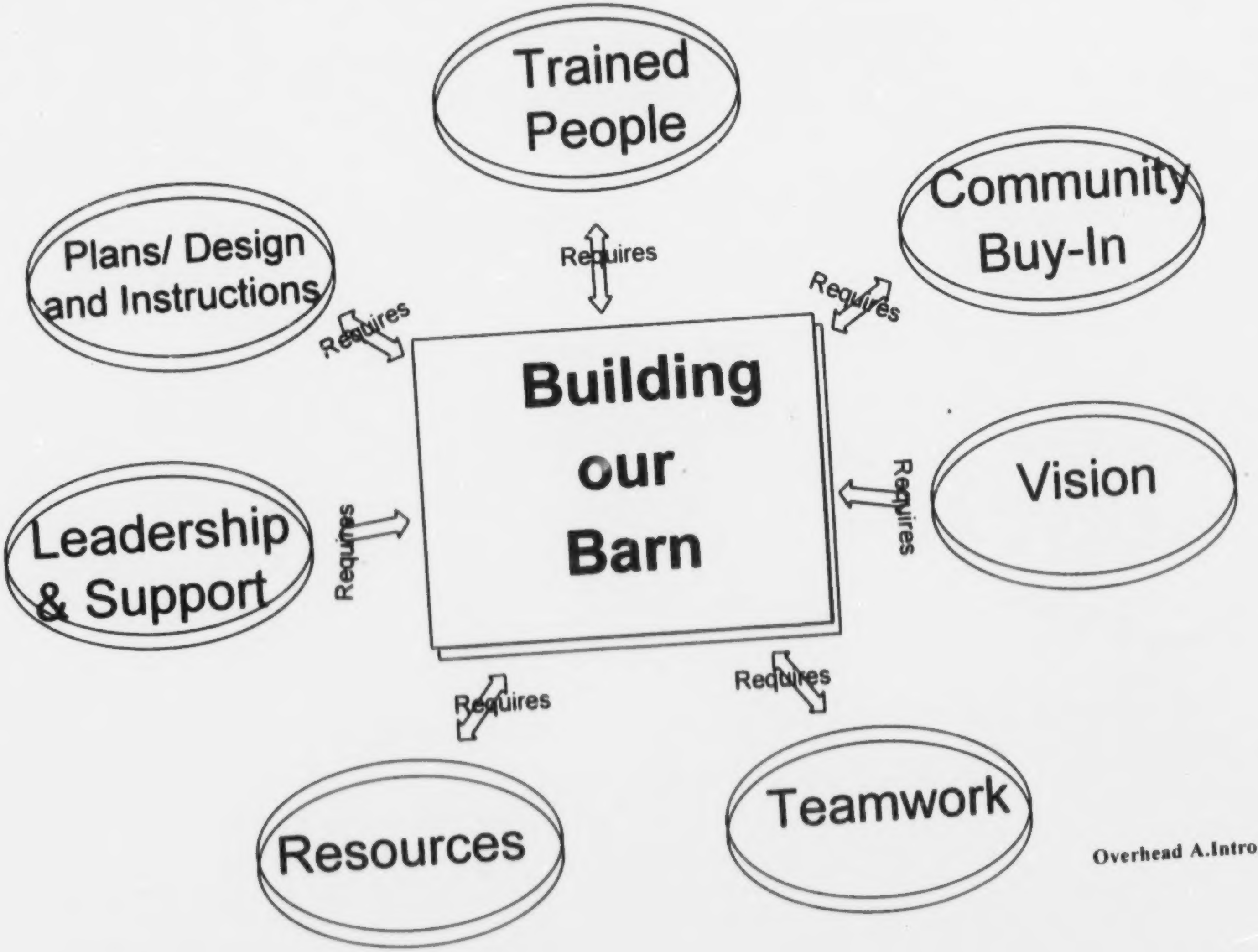
Overhead AIntro.2,O7

Was the barn raising reflective of a working system?



What were the components of this system?

Overhead AIntro.2,O8



Building our Barn

Requires

Trained People

Requires

Community Buy-In

Requires

Vision

Requires

Teamwork

Requires

Resources

Requires

Leadership & Support

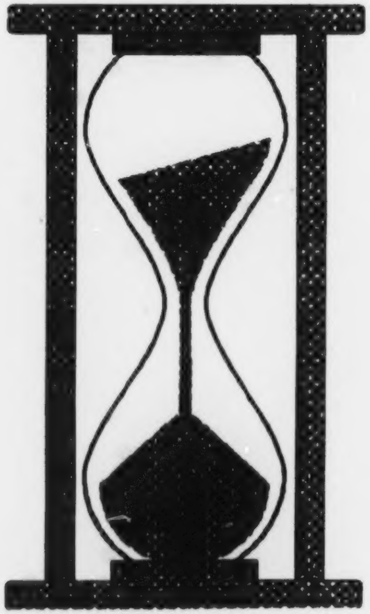
Requires

Plans/ Design and Instructions

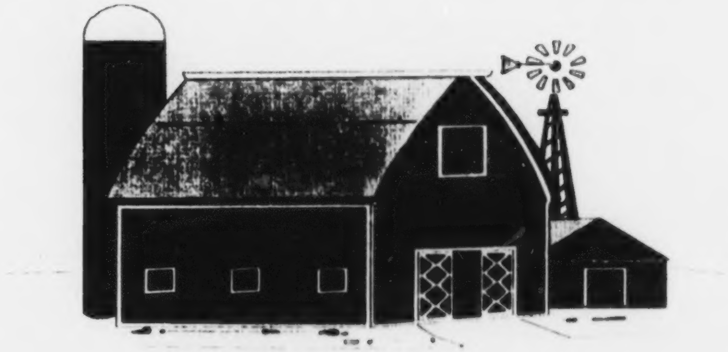
Overhead A.Intro.2,O9

RESOURCES

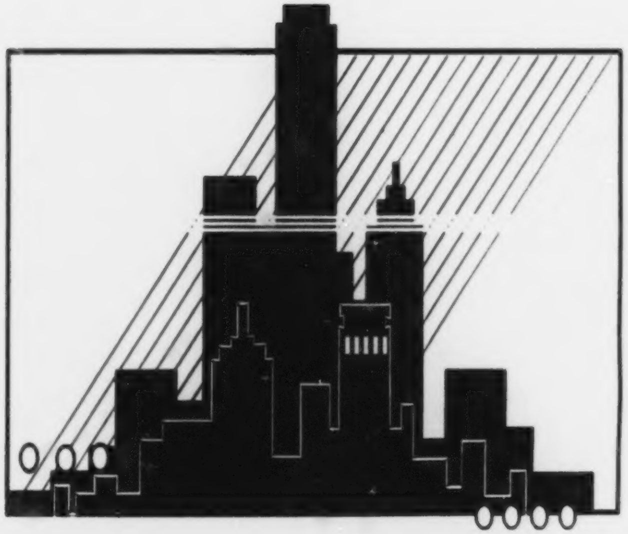
PLAN/DESIGN
AND
INSTRUCTIONS



Overhead AIntro.2,O11

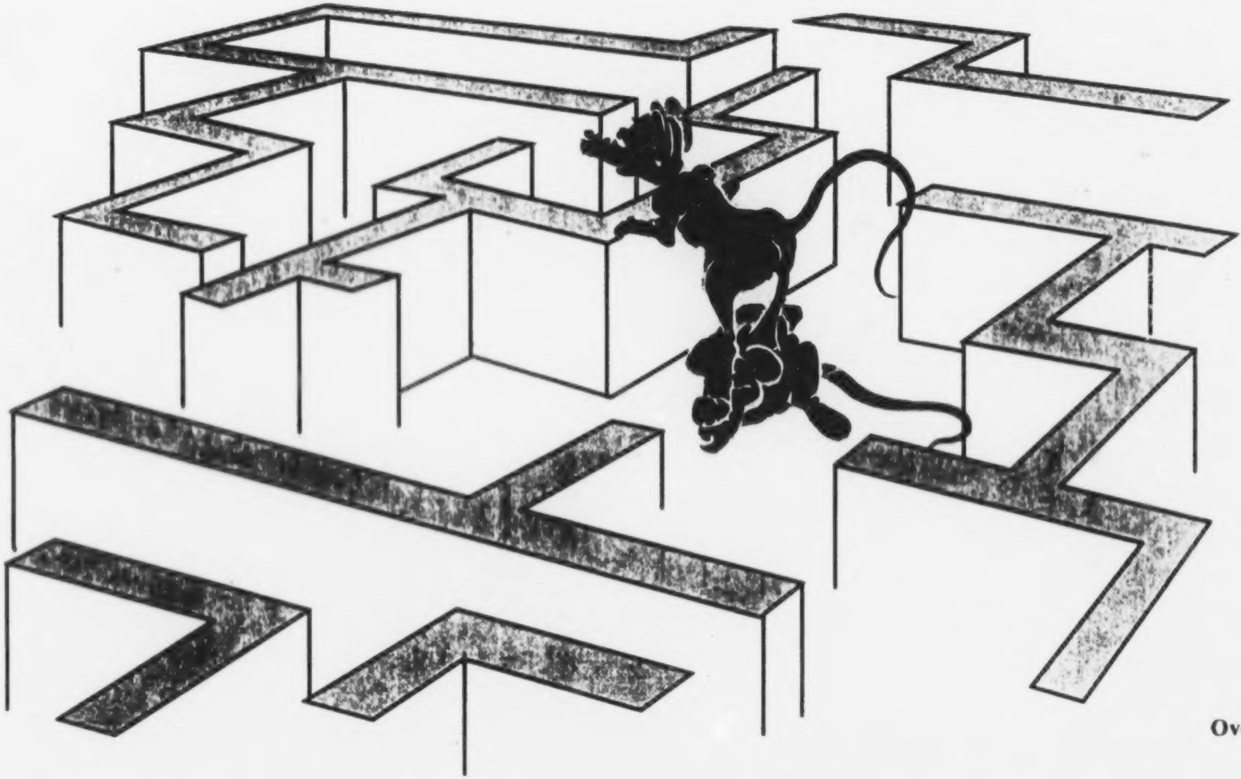


COMMUNITY BUY-IN



Overhead AIntro.2,O12

TEAMWORK



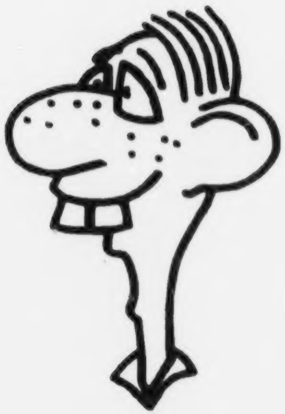
Overhead AIntro.2,O13

LEADERSHIP AND SUPPORT

Overhead AIntro.2,O14

TRAINED PEOPLE

- LEADER
- MATERIAL HANDLERS
- CARPENTERS
- TRAINEES
- FOOD SERVERS

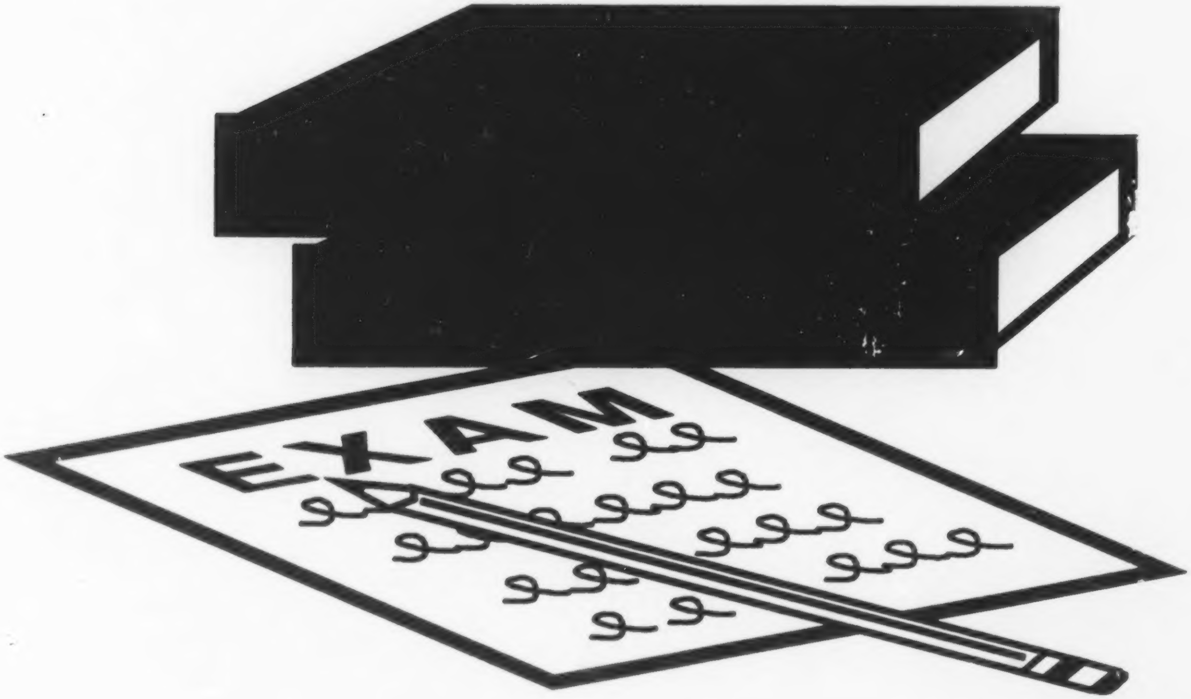


Overhead AIntro.2,O15

A SHARED VISION

Overhead AIntro.2,O16

PLANNING 'QUALITY' GRADE LEVEL OR SCHOOL ASSESSMENT CHANGES

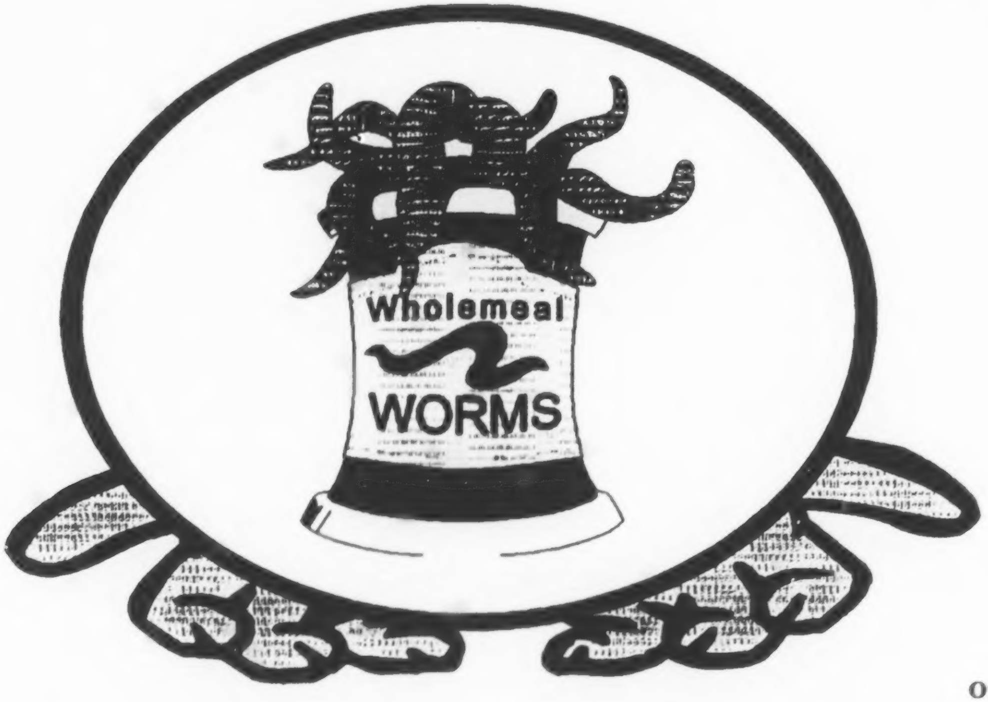


Overhead AIntro.2,O17

PURPOSES

Overhead AIntro.2,O18

A SHARED VISION


Wholemeal
WORMS

Overhead AIntro.2,O19

RESOURCES

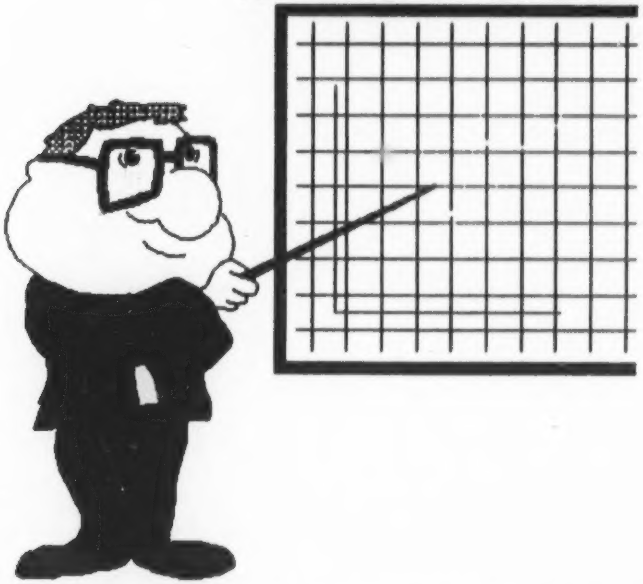
Overhead AIntro.2,O20

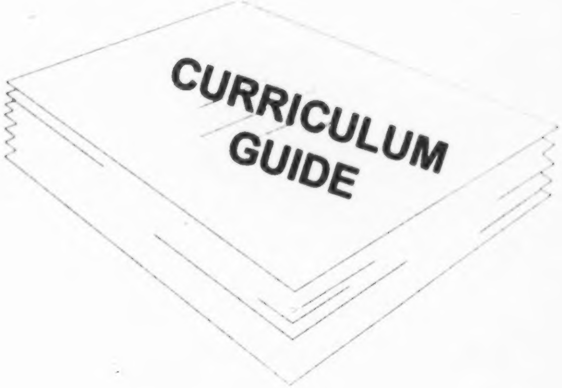
STUDENT & FACULTY BUY-IN

NOW WHAT DO I HAVE TO DO?

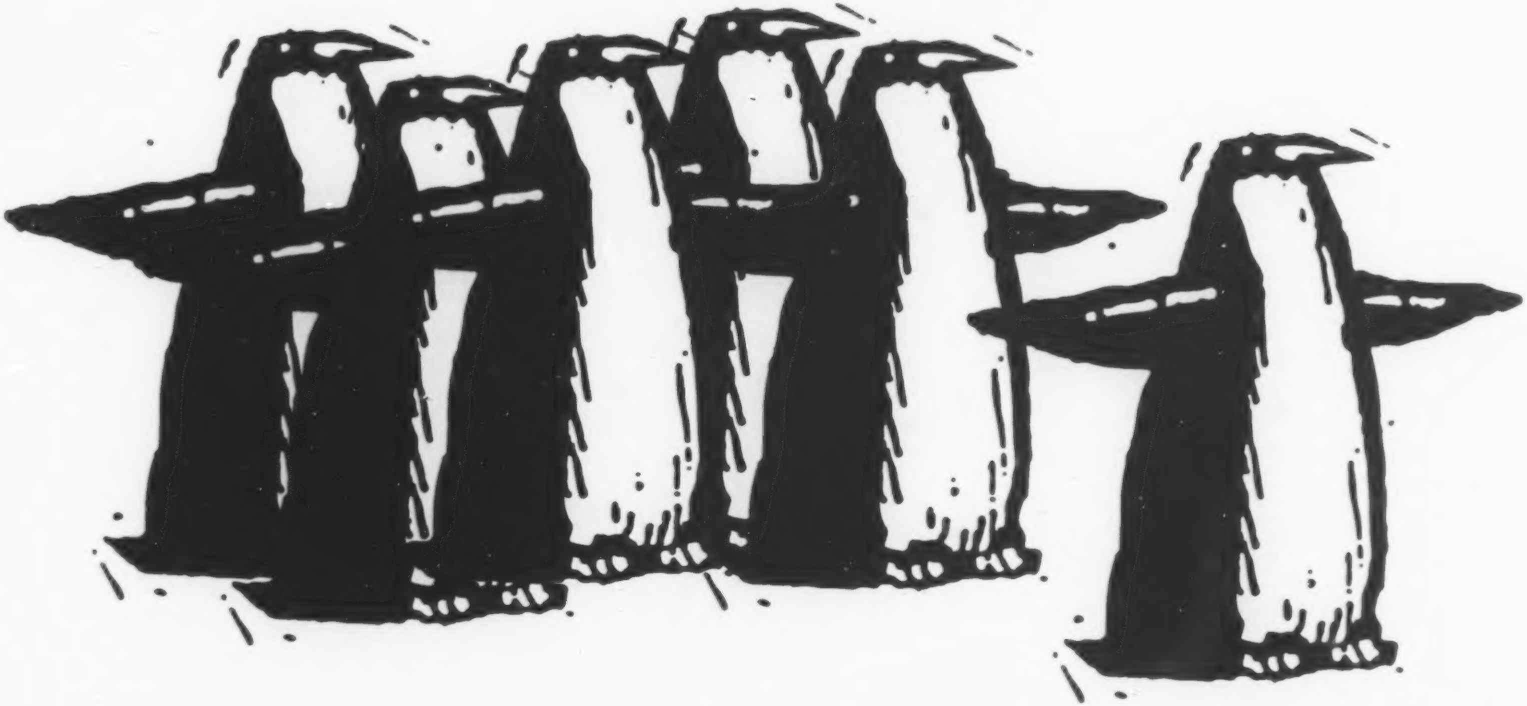
Overhead AIntro.2,O21

CURRICULUM GUIDES AND INSTRUCTION STRATEGIES





LEADERSHIP



Overhead AIntro.2,O23

SUPPORT

- **-Administrative**
- **-Student**
- **-Teacher**
- **-Parent**

Overhead AIntro.2,O24

TRAINED PEOPLE

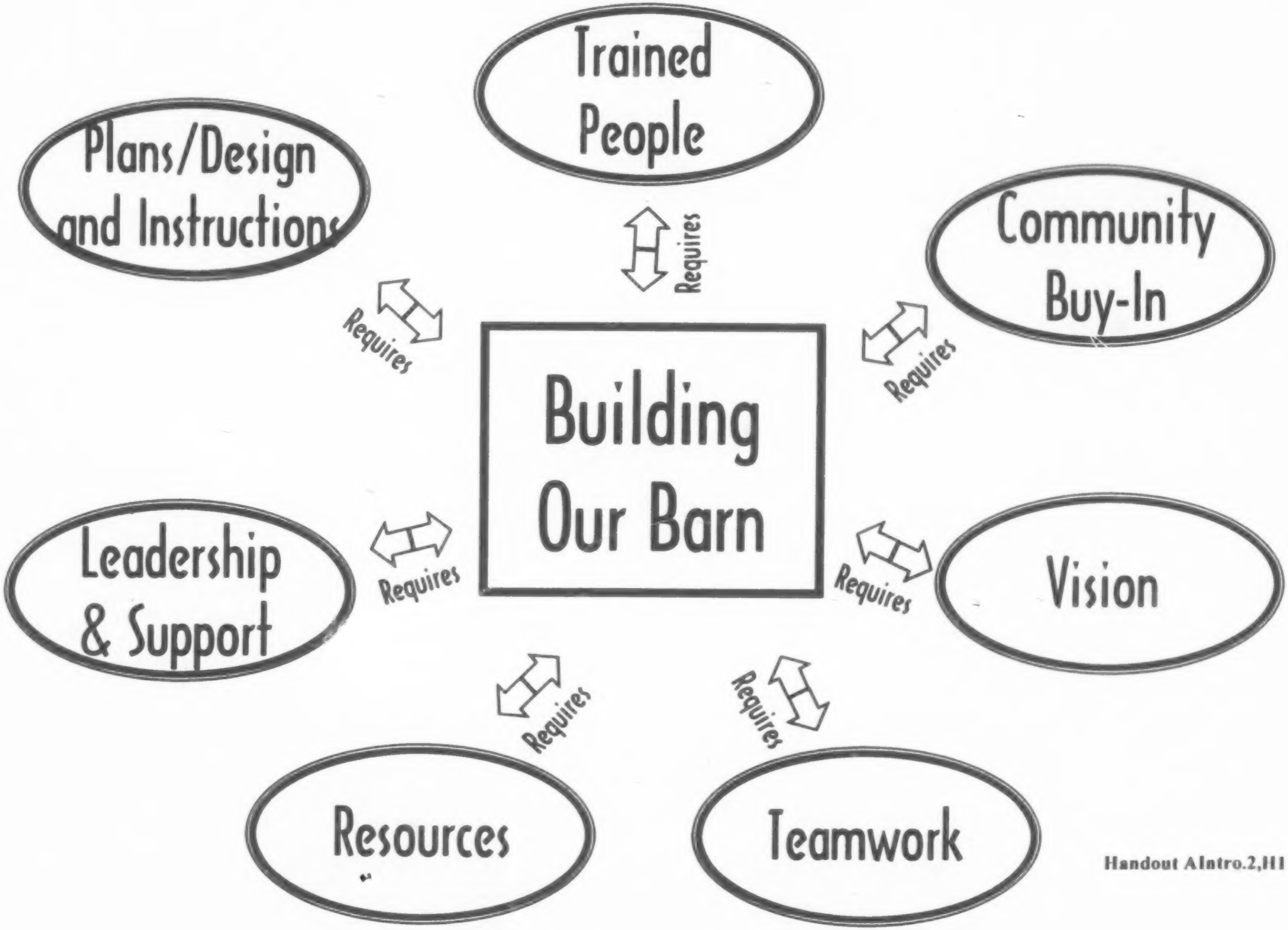
- -STAFF DEVELOPMENT OPPORTUNITIES
- -TRAINER NETWORK
- -TIME TO WORK IN TEAMS

Overhead AIntro.2,O25

SUMMARY

- Planning and reflection time is essential
- Take baby steps
- Change involves a sequence of concerns about the innovation
- Spend your time with people who will lead the way for others
- Promote an environment of continuous growth through feedback spirals

Overhead AIntro.2,O26



Building Our Barn

Trained People

Community Buy-In

Vision

Teamwork

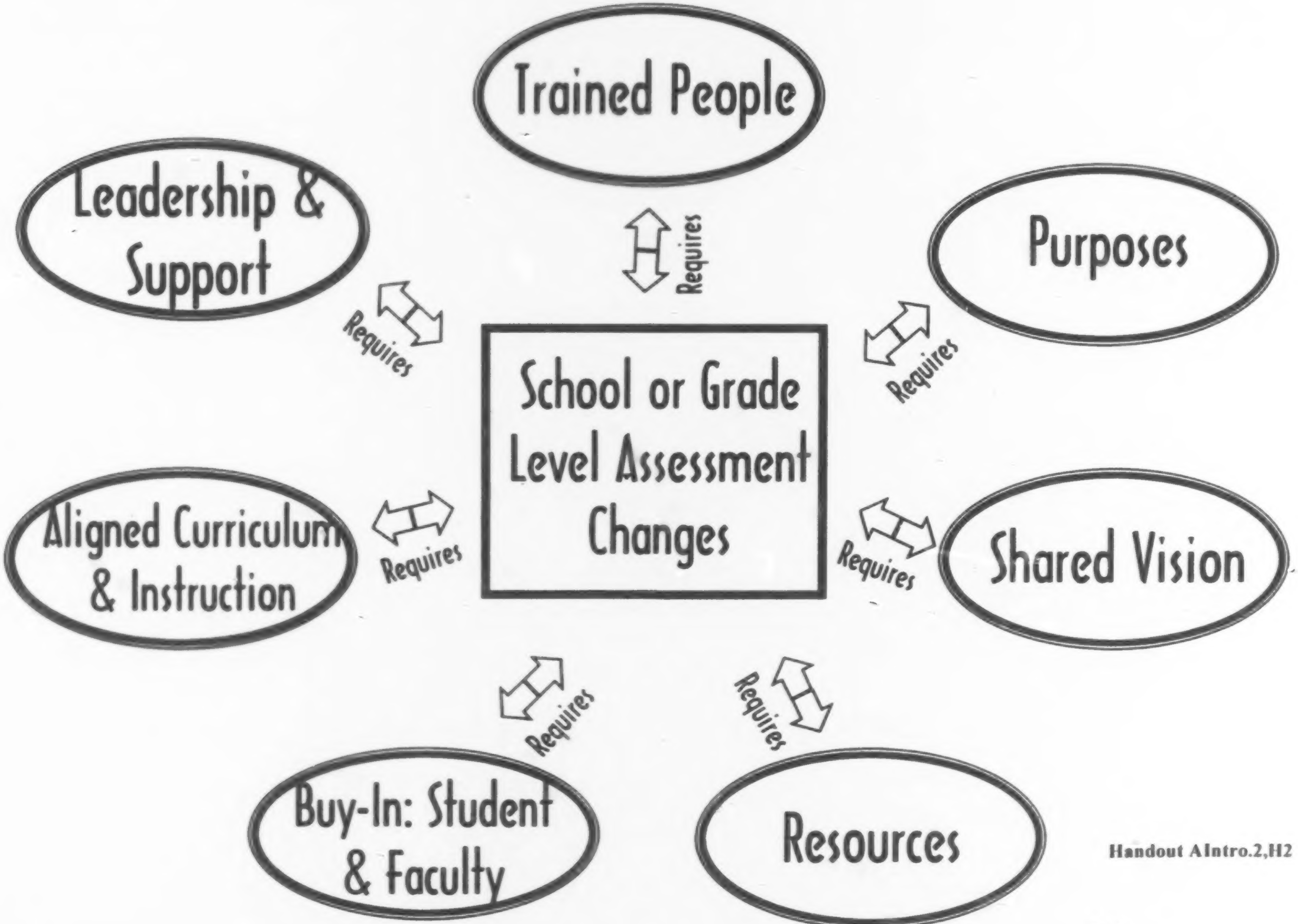
Resources

Leadership & Support

Plans/Design and Instructions

Requires

Handout AIntro.2,H1



School or Grade Level Assessment Changes

Trained People

Purposes

Shared Vision

Resources

Buy-In: Student & Faculty

Aligned Curriculum & Instruction

Leadership & Support

Requires

Requires

Requires

Requires

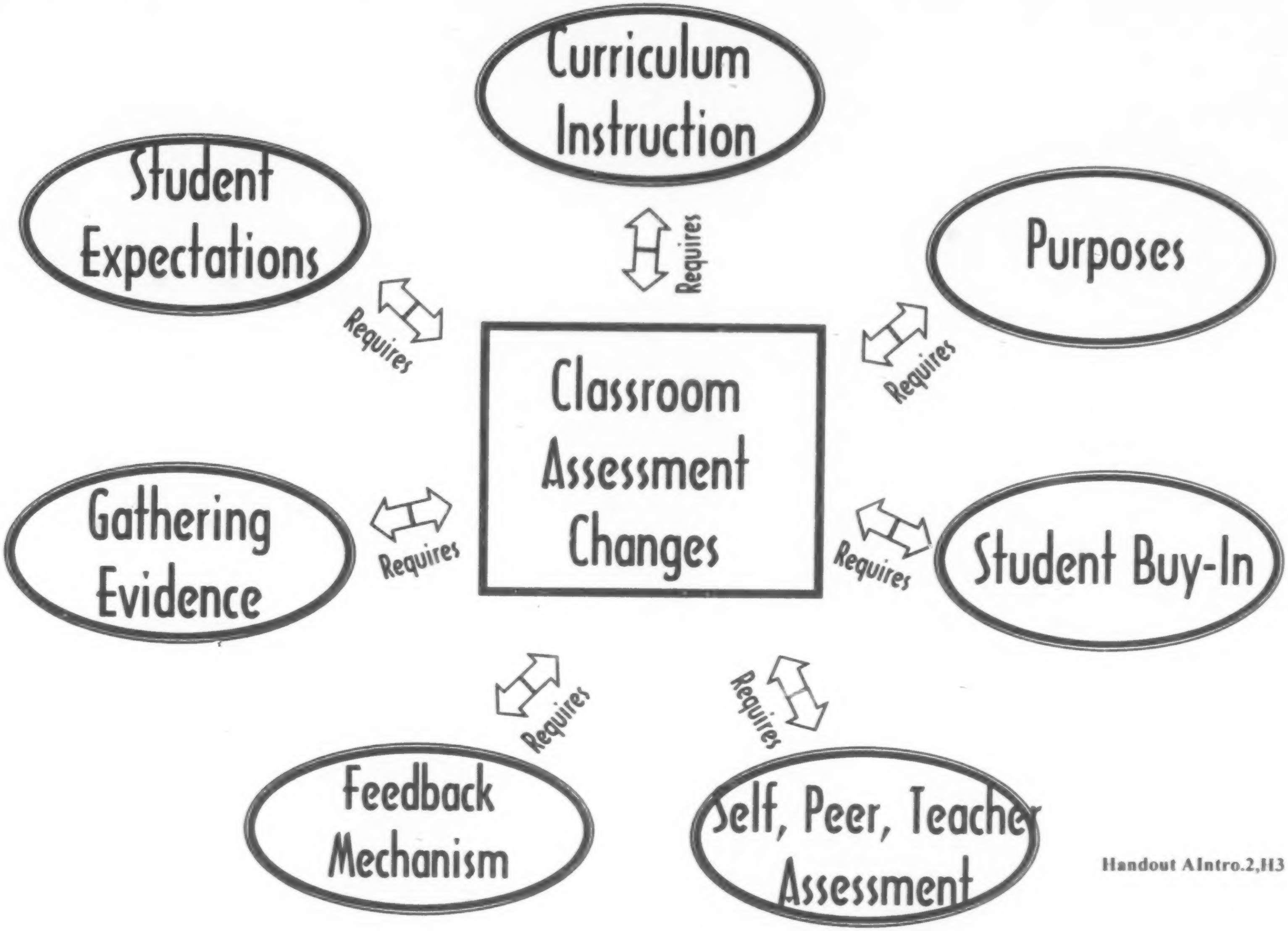
Requires

Requires

Requires

Handout AIntro.2,H2

93



Classroom Assessment Changes

Requires

Curriculum Instruction

Purposes

Student Buy-In

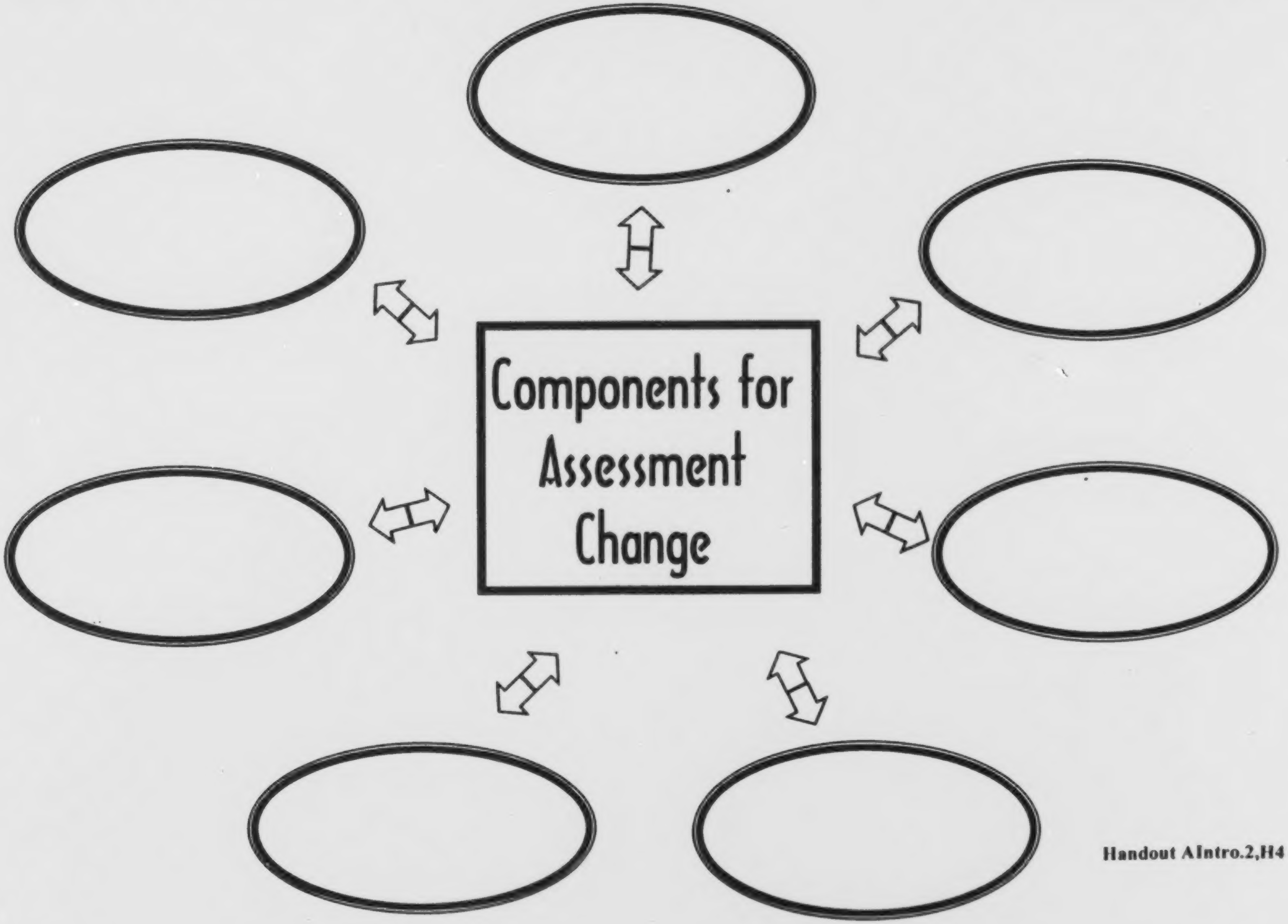
Self, Peer, Teacher Assessment

Feedback Mechanism

Gathering Evidence

Student Expectations

Handout AIntro.2,H3


Components for Assessment Change

Handout AIntro.2,H4

Assessment Action Planning Sheet
(Creating an Assessment Vision for Your School or Grade Level)

What student learning targets do we want our school or grade level to focus on in assessment and why? What will we use the results for? Which students will be assessed?

The Component: ______________________________ *(From our wheel-like resources. Each component requires a separate sheet.)*

Questions to ASK about the component	*Notes from Discussion*
Who are the people responsible for helping create the changes needed? Who do we need to get on board?	
What has been done? How much time did it take?	
What needs to be done? How much time will it take?	
What are the major concerns or barriers for this group?	
What would be the major concerns or barriers for others?	
How will we know if we are successful?	
Do we need major work on this component and what are the group's recommendations?	

Handout AIntro.2,H5

Assessment Action Planning Sheet—(Creating an Assessment Vision for My Classroom)

What is the purpose I want to accomplish with an assessment? For example: monitoring student progress, diagnosing student needs, reporting achievement to others, student self-assessment, etc.

What student learning targets do I want to assess?

The assessment method or process I plan to explore in my classroom and why it serves the purpose and matches the learning targets:

Using the wheel components for Classroom Level Assessment Planning, discuss and respond to the following:

How will the curriculum and instruction be aligned?

What kind of feedback will be given to students to help them improve?

Will this assessment be given a grade? How will the grade be determined? How will you gather information?

How will you communicate student expectations?

How will you promote student ownership (buy-in)?

Will you utilize a self-peer-teacher assessment process?

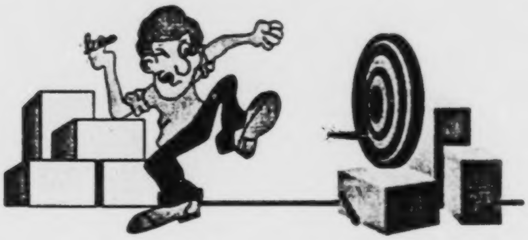
What other components of your assessment method or process needs to be addressed?

Is this assessment method or process worth the time and effort?

Handout AIntro.2,H6

Chapter 1

Standards-Based Assessment—Nurturing Learning



What's in This Chapter?

Think back to the tests you took in school. What were they like? You probably remember some tests as interesting and meaningful experiences, while others you would just as soon forget. The negative ones probably elicited feelings like anxiety, fearfulness, and/or anger. You might even have been nauseated or sleepless. Why was that? What was it about many of the tests we took in school that caused such a reaction? For many of us, our anxiety stemmed from things like being unsure of what would be required; being afraid of poor performance; being afraid to look uninformed; facing the consequences with our parents; having to do something we thought was meaningless; or having the test seem to be so important.

Now think back to the positive tests. What made them so? Most people who recall positive assessment experiences say things such as: "The test was a learning experience," "It was a meaningful challenge that brought a true sense of accomplishment," "It let me know what I knew," or "It was a real life application of skills."

What should assessments be like today? How can we all craft assessments to be more positive and useful experiences for both teachers and their students? What are the changes that are going on in assessment right now and what might we, as teachers, find useful? These are the topics addressed in this chapter.

This chapter can be thought of as the "big picture" chapter—it covers why careful attention to student assessment is crucial and how assessment activities fit into and support other current changes in education. It can also be thought of as the "awareness" chapter—it offers an introduction to basic assessment concepts, explores users' current attitudes toward assessment, and assists users to build a vision of what they want assessment to do for them.

We begin with a look at current thinking about "performance-based" (also called "standards-based") instruction and the role of ongoing assessment of student skills and knowledge to inform practice. Then, we take a quick survey of what it means to assess well. Along the way we visit the reasons that changes in assessment are taking place. The activities at the end of this chapter are designed to explore participants' current familiarity with and level of understanding about assessment, stimulate dialogue about assessment issues, and increase educators' expertise on the meaning and importance of high quality assessment.

Chapter Goals

1. Establish background knowledge about assessment
2. Visit the notion of standards-based education and the role of assessment in the standards process
3. Increase awareness of the principles of good assessment
4. Lay the groundwork for good assessment as a tool for educational improvement
5. Agree on a common language of assessment terms

Chapter Content

A. Readings

Assessment—Where the Rubber Meets the Road in Standards-Based Education ....Page 5

This section introduces the idea that high quality assessment information is necessary for educational improvement and discusses the reasons assessment is changing right now.

Summary of Rationale for Changes in Assessment ....Page 13

Keys to Quality Student Assessment ....Page 14

Since assessment is such an essential part of good instruction, educators need to ensure that their assessments are of high quality. This section discusses characteristics of quality.

Summary of Steps to Quality ....Page 30

Chapter Summary ....Page 32

B. Activities

Activity 1.1 *Changing Assessment Practices—What Difference Does it Make for Students?*

This activity stimulates thinking and discussion among teachers, parents, and diverse communities of learners about why assessment practices are changing. It is a good introductory activity. Time: 20-30 minutes

Activity 1.2 *Clear Targets—What Types Are These?*

In this activity, participants consider different types of learning targets for students, having a good mix of targets, and ensuring that learning targets are crystal clear. It is an intermediate level activity. Time: 40-60 minutes

Activity 1.3 *Post-it Notes™*

This activity reviews assessment terminology and serves as a mixer. It uses the glossary in **Appendix E**, and is a good introductory activity. Time: 20 minutes

Activity 1.4 *Seeing Wholes*

This activity emphasizes the connection between assessment, curriculum, and instruction. It is a good introductory activity. Time: 20 minutes

Activity 1.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore both the meaning of quality with respect to alternative assessment and why we all should care about quality. It is designed for educators at an intermediate level of study about assessment. Time: 75 minutes

Activity 1.6 *A Comparison of Multiple-Choice and Alternative Assessment*

Participants compare a multiple-choice test to an alternative assessment that attempts to measure the same skills, and discuss the advantages and disadvantages of each approach. This is an introductory level activity which uses assessment samples in **Appendix A**. Time: 30-60 minutes

Activity 1.7 *Target-Method Match*

This activity introduces assessment methods, and gives participants practice in matching methods to learning targets. It is designed for educators at an intermediate level of study about assessment. Time: 60-90 minutes

Activity 1.8 *Sam's Story: Comprehensive Assessment*

This activity illustrates the need for multiple measures of student achievement. It is an introductory level activity. Time: 45 minutes

Activity 1.9 *Going to School*

Part A of this activity demonstrates the importance of performance criteria; Part B illustrates different types of performance criteria, and Part C discusses the advantages and disadvantages of various types of rubrics. The activity uses assessment samples in **Appendix A**. Part A is an intermediate activity; Parts B & C are advanced. Time: 60 to 90 minutes

Activity 1.10 *Clear Targets and Appropriate methods—The View From the Classroom*

Participants self-evaluate the extent to which their instruction is focused on clear targets and the degree of conscious matching they perform between targets and assessment methods. Self-rating rubrics are provided. This activity is intermediate in difficulty. Time: 30-45 minutes

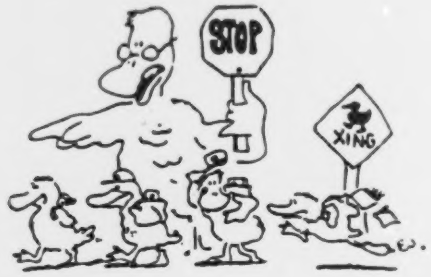
Activity 1.11 *Assessment Standards*

This activity provides an opportunity for participants to look at assessment standards in various content areas and related examples of assessment questions. It is an intermediate level activity. Time: 90 minutes

Activity 1.12 *Assessment Principles*

Participants examine the beliefs that influence their decisions about student assessment and explore equity issues in assessment. It is an introductory activity. Time: 20 minutes

Standards-Based Assessment—Nurturing Learning Readings



Assessment—Where the Rubber Meets the Road in Standards-Based Education

Why is there such strong belief in the power of assessment to motivate changes in teaching and learning? The logic appears relatively simple:

- Assessments are the final word in defining what educators and the community want students to know and be able to do. Regardless of what anyone says in their standards documents (or in course outcomes), the assessments define what was really meant. So, assessments communicate the standards to which school systems, schools, teachers, and students aspire;

- These standards provide focus and direction for teaching and learning;

- Results from assessments support important insights on the nature, strengths, and weaknesses of student progress relative to the standards;

- Educators and students can use this feedback to understand and direct their attention to improving relevant aspects of student learning;

- Participating in the assessment itself can be a powerful episode of learning for both teachers and students; such participation can directly affect instruction and student achievement.

The power of assessment to influence teaching and learning isn't always positive. Just consider some of the negative consequences from past over-reliance on standardized multiple-choice tests. Under pressure to help students do well on such tests, teachers and administrators tended to focus their efforts on test content, mimic the tests' multiple-choice formats in classroom curriculum, and devote more and more time to preparing students to do well on the tests. The net effect was a narrowing of the curriculum to the basic skills assessed and a neglect of complex thinking skills and other subject areas which were not assessed.

This is why assessment is such an important part of the standards-based reform efforts which are going on across the country. Nearly every state and many, many local districts and schools are engaged in both setting rigorous standards that define what students should know and be able to do for future success and developing assessment systems to mirror these high standards. Through these assessments, states, districts and schools are trying to support and promote the attainment of their standards.

For a more personal, classroom-based example, consider these thoughts by Francine Staytor and Peter Johnston: *What we choose to evaluate and how we choose to evaluate delivers powerful messages to students about those things we value. Students view their learning and their sense of worth through the lens we help them construct unless they cannot bear to look through it.* The authors then give many examples of these messages, such as focusing on errors rather than on strengths, focusing on conventions rather than meaning, and projecting the image that the teacher is the only one that has the knowledge and ownership of the correct response.

📖 Reference Box:

For further reading on the importance of assessment in education see: Joan Herman, *Assessing New Assessments: How Do They Measure Up?*, 1990, Los Angeles: UCLA Graduate School of Education, (310) 206-1532.

For further reading on the effect of multiple-choice tests on instruction see: H.D. Corbett and B.L. Wilson, *Testing, Reform, and Rebellion,* 1991, Norwood, NJ: Ablex Publishing.

For further reading on the ways that classroom assessment can affect students see: Francine Stayter and Peter Johnston, "Evaluating the Teaching and Learning of Literacy," in Timothy Shanahan, (Ed.), *Reading and Writing Together: New Perspectives for the Classroom,* 1990, Christopher-Gordon Publishers.

Learning From the Past

If what one gets is what one assesses, then we all need to make sure that our assessments truly reflect our standards for student accomplishment, clearly communicate the important things we want students to know and be able to do, and encourage reflective teaching. Many of the current trends in assessment (such as more emphasis on alternative forms of assessment—performance assessment, senior projects, portfolios) are attempts to do just this.

Many people believe that the characteristics of alternative assessment reflect current understanding of how students learn best, and are more able to assess the kinds of student learnings that we increasingly feel are important to be successful in the 21st century. These abilities include:

- Applying knowledge by creating, designing, producing, and performing
- Tapping more complex thinking and problem solving skills
- Requiring real life applications
- Requiring real people, rather than machines, to judge the quality of student work

If educators have learned anything from the past 60 years of testing, it is this: If we want our assessments to have positive consequences, we first have to have a clear vision of what we want our assessments to accomplish and then design the assessment to do it. This sequence is the essence of quality and is a major theme that will be visited many times throughout *Toolkit98.*

Toolkit98, and the Laboratory Network Program that led to its creation, recognizes the important role assessment plays in the educational improvement process. High quality assessment targeted at important goals for students supports planning for educational improvement. On the other hand, assessment that focuses on the wrong goals, models outmoded instructional practice, or is used for the same old purposes can impede reform.

For all these reasons, educators and educational policymakers are working to change the assessment of student achievement to reflect changes in the expectations and standards for students, in the methodologies and purposes of assessment, and in the instruction that assessment is intended to support. In *Toolkit98*, we refer to this evolving type of assessment as *alternative assessment.* As noted in the *Introduction*, the definition we rely upon is:

> *Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.*

Also, as noted in the *Introduction*, even though *Toolkit98* emphasizes alternative assessment, this does not imply that only alternative assessments are worthwhile and all previous assessment efforts are worthless. The key to successful decision making is to provide a comprehensive assessment of important student goals, using the most appropriate combination of assessment tools to do so.*

---

* For this reason, multiple sources of assessment information about student achievement, including assessments that will allow for the evaluation of complex skills, are important. No single type of assessment will provide all

Rationale for Alternative Assessment

Why do educators need alternative assessments? There are a variety of reasons detailed in this section.

Changes in Valued Student Outcomes. Standards and assessments of student achievement are changing because students face a world that will demand new skills and abilities. Society's knowledge base has grown exponentially over the last few decades, and continues to do so. With the volume of information doubling every three years, there are too many facts to learn.

> *Given this pace, no one individual can be expected to keep up with the information flow in a single discipline, much less across disciplines. Such a knowledge explosion makes futile most attempts to have students memorize and regurgitate large bodies of facts.*
>
> *Economic trends also push us away from a fact-based curriculum. The shift from a manufacturing- to an information- and service-based economy requires that individuals have skills in accessing and using information and in working with people. These changes in the workforce and in the pace and complexity of modern life suggest that people will need to be flexible, to shift jobs frequently, and to adapt to change. To prepare students for success in the future, schools must emphasize how to apply rather than just acquire information.* (Herman, et. al, 1992, p. 14)

🕮 Reference Box:

Joan Herman, Pam Aschbacher, and Lynn Winters, *A Practical Guide to Alternative Assessment*, 1992, Alexandria, VA: Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

In the future, students will need to know how to access the information they need and apply it to real-life situations. Students will face many situations where there will be no clear-cut right answer, but where, instead, they will need to analyze the situation and apply their knowledge and skills to find a solution that will work. Yes, knowledge and facts will still be important, but they simply are not sufficient to prepare students for future success. Today's assessments therefore need to measure not only the basics, but also a student's ability to think critically, analyze, and make inferences—skills found in the "content standards" adopted recently by a variety of national commissions. In these efforts, expectations for students reflect an increasing emphasis on critical thinking, problem solving, the ability to monitor one's own performance, the efficiency with which tasks are accomplished, group collaboration, and communication skills. Such abilities are difficult to measure in multiple-choice format; practitioners need alternatives.

---

the information that teachers need. In this new era of assessment, all of the assessment techniques and strategies that are available will be needed.

In a thinking curriculum, students develop an in-depth understanding of the essential concepts and processes for dealing with those concepts, similar to the approach taken by experts in tackling their tasks. For example, students use original sources to construct historical accounts; they design experiments to answer their questions about natural phenomena; they use mathematics to model real-world events and systems; and they write for real audiences. (Herman et al., 1992, p.17)

Reference Box:

For a summary and compilation of content standards across a variety of disciplines see: John Kendall and Robert Marzano, *Content Knowledge: A Compendium of Standards and Benchmarks for K-12 Education*, 1997, McREL, (303) 337-0990, http://www.mcrel.org/

In this compendium, the authors note, "There is clearly not a consensus across groups as to what form 'standards' should take or how they should be used. The result is that the character, scope, and level of detail provided in standards often vary significantly from one subject area to another."

For a history of national standards development see: Robert Marzano and John Kendall, *Designing Standards-Based Districts, States, Schools and Classrooms*, 1996, Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

For skill standards used in career-technical education visit: http://www.nssb.org/ossp.html or click on the STW Links button at http://www.stw.ed.gov/

For those interested in standards from various states, see: http://putwest.boces.org/Standards.html#Section3

Changes in Notions of Competence. Notions of competence and expertise are also becoming more refined. For example, the definition of what it means to be a good reader has changed over time. Now it is not enough just to be able to convert symbols on the page to sound. To read at the levels necessary to be functionally literate in tomorrow's world, students need to be able to draw inferences, relate current information to past information, have a variety of efficient reading strategies, know when they are not understanding, and so forth. These skills are very difficult to measure in fixed-response format because it is not just a matter of testing separate skills, but also a matter of assessing a student's ability to know how and when to use a variety of skills collectively to accomplish a reading goal.

Similarly, traditional science courses required students to learn vocabulary and memorize basic scientific principles. Today, people recognize that memorizing formulas does not necessarily result in "expertise" and may, in fact, be detrimental to the development of expertise. Students must understand and be able to apply those scientific principles to real-life problems. More science classrooms today look like laboratories where students are

BEST COPY AVAILABLE

learning by doing. Science assessments are beginning to change from mere vocabulary tests to assessments that enable students to demonstrate their conceptual understanding of the material.

Changes in Teaching and Learning That Are Communicated by Assessment. Teachers, parents, and students develop an understanding of what is valued in education by virtue of what is assessed and how it is assessed. For example, when multiple-choice tests are used exclusively, it gives the impression that there is only *one* right answer to every question and that there is *always* a right answer. Kober (1993) observes the following messages from many past multiple-choice, norm-referenced, traditional science tests:

a. Because they have a single right or wrong answer, they reinforce the misleading conception of science as a static body of facts

b. Because their results rank individual student performance against that of a larger group, they perpetuate the notion that only a few students—the top scorers—are smart enough to pursue science

c. Because they sample a breadth of content in a superficial and unconnected way, traditional tests actually reward instruction that drills students on low level facts and vocabulary recognition*

Such "traditional" assessment goes along with a view that teachers are providers of knowledge and students are passive recipients of that knowledge. Research says, however, that different instructional approaches are required if students are to acquire the thinking skills needed for tomorrow's world. Current research indicates that good instruction engages students actively in the learning process.

🕮 Reference Box:

For more information on current knowledge about instruction and the learning process see: N. Kober, *EdTalk: What We Know About Science Teaching and Learning*. National Education Knowledge Industry Association—NEKIA, (formerly Council for Educational Development and Research), (202) 429-5101.

> *Current evidence makes it clear that instruction emphasizing structured drill and practice on isolated facts and skills does students a major disservice. Insisting that students demonstrate a certain level of arithmetic mastery before being allowed to enroll in algebra or that they learn how to write a good paragraph before tackling an essay are examples of this discrete skills approach. Such learning out of context makes it more difficult to organize and remember the information being presented. Applying taught skills later when solving real-world problems also becomes more difficult. Students who have trouble mastering*

---

* We are not saying that *all* multiple-choice tests *have to* fall into these traps; in fact, there are many attempts to make multiple-choice tests that assess knowledge in a more connected fashion.

> *decontextualized 'basics' are often put in remedial classes or groups and are not given the opportunity to tackle complex and meaningful tasks.* (Herman et al; 1992, p.15)

Assessment must change to support new directions in instruction. What kind of reinforcement do teachers receive for making necessary changes in classroom practice if the outcomes assessed by the tests do not cover the full range of desired outcomes and the procedures don't correspond to what is known about sound instruction?

Changes in the Purposes of Assessment. Assessment in the past was most frequently used for sorting students: grading, selection into special programs, assignment to instructional groups, identification of the highest and lowest performing students, and so forth. Different skills were expected of students depending on their ability.

In today's information age, *all* students need to meet high standards in order to succeed in the world. Educational assessment can no longer primarily play a sorting role. Rather, assessment must help identify the assets of students on which effective educational programs can be built. It must also identify the characteristics that are likely to interfere with the student's learning so that the school or college may help overcome these difficulties.

Today, knowing that one student ranks higher than another is less important than knowing how both students' performances compare with our ultimate goals for performance. It is important for assessments to describe student performance well enough so that students and teachers know how students are progressing toward agreed upon goals. The standard is no longer "doing better than the rest," but rather meeting a previously agreed upon, rigorous level of knowledge and skills. In short, the major purposes of assessment are to improve student learning and inform instruction.

Changes in the Importance of Authenticity. Multiple-choice test results are "stand-ins" or proxies for the actual performance of interest. For example, educators could use a multiple-choice test to assess knowledge about how to give an oral presentation. If the student does well on the knowledge test, we might infer that the student could actually *do* a good oral presentation. Performance on the multiple-choice test is therefore a proxy for the actual performance. However, although students who do well on multiple-choice tests tend to perform better as well, the correlation does not hold for all students or all occasions. To use a multiple-choice test as the exclusive measure of a student's ability to give an oral presentation ignores the need to assess the application of that understanding in an actual performance task and is likely to constrain instruction to what is on the multiple-choice test. Now, one might think that giving an oral presentation is a silly example—of course one would need to have a student actually *do* an oral presentation. But the same logic applies to laboratory skills, writing, and a host of other complex skills we want students to master.

Modern theories of teaching and learning have demonstrated that students know, understand, and retain more when they learn in the context of real-life situations. They can also demonstrate the depth of that understanding when the task they are asked to perform mirrors a real-life situation. Educators recognize the need to give students more realistic tasks to do on assessments because they want to know if the students can apply their knowledge to more real-life situations, and because they don't know whether artificial situations really elicit the full range of what students are able to do.

Changes in the Recognition of the Critical Role of Teachers in Assessment. More attention is being paid to helping teachers prepare themselves for classroom assessment and classroom uses of alternative assessment for a number of reasons:

1. Assessment is a frequently occurring activity in the classroom, yet teachers rarely are given the opportunity to develop the skills they need to assess well and to use assessment information to improve instruction.
2. Teachers are at the heart of educational improvement. It is they who will administer and interpret the results of alternative assessments. If teachers don't feel a need for this information, or don't see how it can be used to improve student learning, their resistance may result in the same contamination of assessment results that occurs with more traditional large-scale assessments.
3. Alternative assessment is not magic. Just having an alternative assessment in place won't automatically improve student achievement. Teachers will need to understand principles of assessment in order to ensure that new types of assessments are not misused or under-utilized. We need to build in those features that will make such assessments useful in the classroom, and then help teachers understand how to use them to improve the learning of their students. If we don't, the supposed impact of our alternative assessments may not materialize.
4. Changes that matter occur in classrooms, in the day-to-day interactions between teachers and students. Properly designed alternative assessments can support changes in these interactions. Our knowledge of how students learn is increasing. Our assessments should reflect our new knowledge about teaching and learning.

With the increasing demands placed on teachers, support given through time and professional development is critical to the success of any educational improvement effort. *Toolkit98* is designed to provide one resource to help ensure that teachers will receive the support they need to use assessment in the service of instructional improvement.

Related *Toolkit98* Chapters and Activities:

Activities that relate to the rationale for changes in assessment are: **Activity 1.1**—*Changing Assessment Practices: What Difference Does it Make for Students?;* **Activity 1.4**—*Seeing Wholes*; **Activity 1.6**—*A Comparison of Multiple-Choice and Alternative Assessment*; and **Activity 1.12**—*Assessment Principles.*

Summary

The rationale for alternative assessment and a description of the way it fits into current efforts to improve student achievement is summarized very nicely by the following quotation:

> *The area of achievement assessment has been undergoing major changes during the past few years. A shift has taken place from what some call a "culture of testing" to a "culture of assessment." A strong emphasis is put on integrating assessment and instruction, on assessing process rather than just products and on evaluating individual progress relative to each student's starting point. The position of the student...has also been changing...to that of an active participant who shares responsibility in the process, practices self-evaluation, reflections, and collaboration and conducts a continuous dialogue with the teachers. The [assessment] task is often interesting, meaningful, authentic and challenging....All these changes are part of a school restructuring process, meant to develop self-motivated and self-regulated learners and intended to make learning a more mindful and meaningful experience which is responsive to individual differences among the learners. This shift reflects an "overall assessment prophecy" which holds that it is no longer possible to consider assessment only as a means of determining which individuals [can adapt] to mainstream educational practice....[Now] rather than requiring individuals to adapt to means of instruction, the desired objective is to adapt the means of instruction to individuals in order to maximize their potential for success....The new assessment alternatives being developed enhance the possibilities for adaptation....* (Birenbaum and Douchy, 1996, p. xiii)

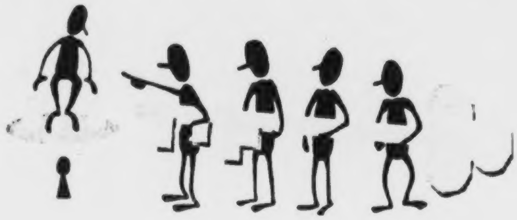
📖 **Reference Box:**

Menucha Birenbaum and Filip Douchy, 1996. *Alternative in Assessment of Achievements, Learning Processes and Prior Knowledge*. Kluwer Academic Publishers, (781) 871-6600.

The Need for Quality

Given the important role assessment plays in education (and educational reform efforts), it behooves everyone to make sure that assessments are of high quality. It's perfectly possible to design poor quality alternative assessments and use the results for the wrong purposes. Therefore, we're going to back up a little and talk about what good assessment looks like in general. We'll consider all forms of assessment, not just alternative forms of assessment.

Keys to Quality Student Assessment



Sound assessments at any level from the classroom to the boardroom[1]:

- **Arise from clear and appropriate student learning targets.** *What* are we, as educators, trying to assess? We all must clearly and completely define achievement expectations, and these must be couched in the best current understanding of the discipline. We can't assess something if we don't have a crystal clear vision of what it is. But, we can have crystal clear learning targets that aren't enduring or essential. So, we also need *appropriate* learning targets for students.

  "Student learning targets" are also called many other things: content standards, benchmarks, learning objectives, outcomes, learning goals, essential academic learning requirements. They all attempt to define clear and appropriate achievement targets for students. The role of assessment then is to align with the targets.

- **Serve a focused and appropriate purpose.** *Why* are these targets being assessed? Who will use the results and what will they be used for? Purpose affects how one assesses. But, one can have focused purposes for assessment that are *poor* purposes (such as assessing for the sole purpose of tracking students), so one also needs to make sure that purposes are *appropriate*.

🕮 **Reference Box:**

For further reading on keys to quality assessment at the classroom level, see: Rick Stiggins, 1997, *Student Centered Classroom Assessment*, pp. 14-17. Prentice-Hall, (201) 236-7000.

For more information on keys to quality assessment at the large-scale level see: Joan Herman, 1996, "Technical Quality Matters," in Robert Blum and Judy Arter (Eds.), *Student Performance Assessment in an Era of Restructuring*, Section I, Article #7. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723. ISBN 0-87120-267-0.

---

[1] These ideas are adapted from the two authors listed in the reference box.

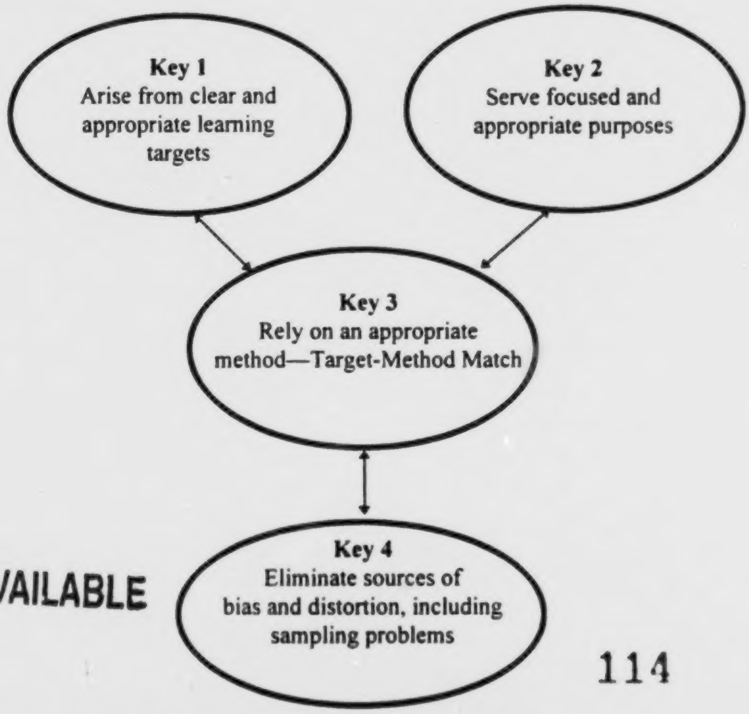
- **Rely on an appropriate method.** *How* will one assess the achievement targets? Will these methods accurately reflect the achievement targets and purposes? When is the best time to use multiple-choice, essay, performance assessments, or portfolios?

- **Eliminate possible sources of bias and distortion.** *How good* is the assessment? Is there anything in the way an achievement target is assessed that masks the true ability of a student or group of students? *How much* will be collected? Can one be confident that results really reflect what a student knows and can do? Technical terms for the notion of reducing bias and distortion are: reliability, validity, fairness, equity, and sampling.

These keys are also shown in **Figure 1.1** below. The rest of this section will add a little more detail to each key. Chapter 3 will expand on these ideas with respect to alternative assessment.

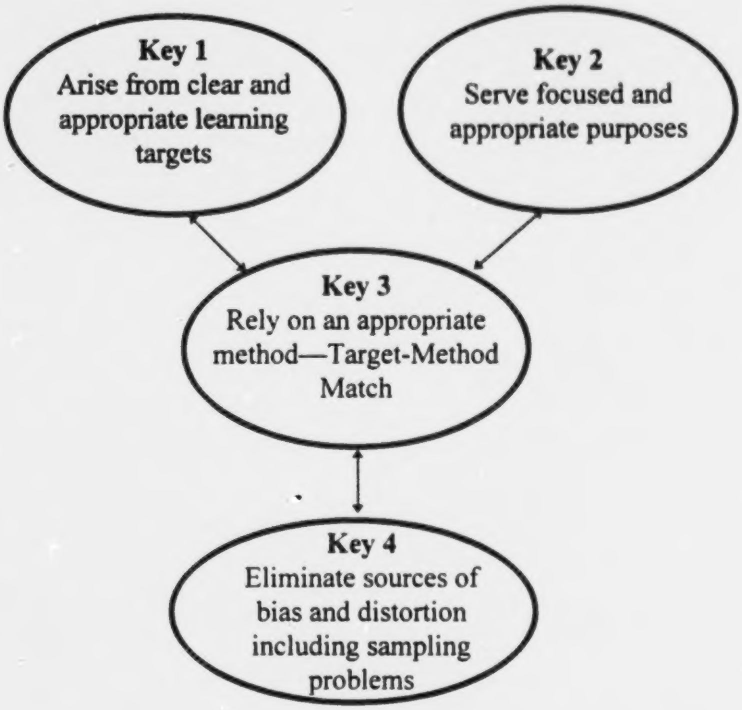
Figure 1.1
Four Keys to Quality Student Assessment

Good quality assessments:


Key 1
Arise from clear and appropriate learning targets
Key 2
Serve focused and appropriate purposes
Key 3
Rely on an appropriate method—Target-Method Match
Key 4
Eliminate sources of bias and distortion, including sampling problems

BEST COPY AVAILABLE

Key 1: Clear and Appropriate Learning Targets


Key 1
Arise from clear and appropriate learning targets
Key 2
Serve focused and appropriate purposes
Key 3
Rely on an appropriate method—Target-Method Match
Key 4
Eliminate sources of bias and distortion including sampling problems

The first key to quality with respect to assessment is to have clear and appropriate learning targets. One can't assess something if one doesn't know what it is they're trying to assess.

For Key 1, we'll tackle three related topics. The first is "content standards." Standards setting activities have blitzed the country over the past eight years and are intended to define the "appropriate" part of the "clear and appropriate student learning targets" equation. So, educators need to know about them. Secondly, we'll address the "clear" part of the equation. Then, finally, we'll present a couple of ways to categorize learning targets, and discuss why we might want to classify them.

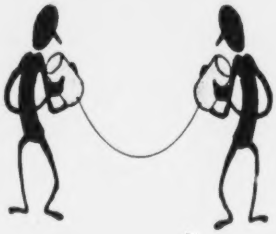
Appropriate Targets/Content Standards. As we've mentioned previously in this chapter, content standards are statements of what should be taught; they specify the "what" of what students should know and be able to do. Content standards come by many names—benchmarks, outcomes, essential academic learning requirements, skills standards, competencies, common curriculum goals, and academic student expectations. Here are some examples:

- Oregon, Grades 6-8 Reading Standard: *Demonstrates inferential comprehension of a variety of printed materials.* Related grade 8 Benchmark: *Identify relationships, images, patterns or symbols and draw conclusions about their meaning.* ("By Grade

Level Common Curriculum Goals, Grades 6-8 Content and Performance Standards," Oregon Department of Education, August, 1996)

- Washington, Writing Standard, Grades 4-10: *The student writes clearly and effectively.* This standard includes the following "components": *develop concept and design, use style appropriate to the audience and purpose (voice, word choice, and sentence fluency),* and *apply writing conventions.* (Essential Academic Learning Requirements, Washington State Commission on Student Learning, January 1997)
- National Standards for Business Education, Secondary: *Demonstrate interpersonal, teamwork, and leadership skills necessary to function in multicultural business settings.* (National Standards for Business Education, National Business Education Association, 1995)

"Nothing new," we hear you thinking, and in a very real sense you're right. The idea of focusing instruction and assessment on that which is most important and enduring is not new. The hard part of it is coming to agreement on what is important and enduring. Have you ever had disagreements with your colleagues on the most important things to emphasize in instruction? Expand this a thousand-fold as a nationwide effort, and you get the picture. Some of the efforts to set standards read like soap operas. But this only underscores the necessity to do it. How can anyone hold students responsible for outcomes for which they disagree on the meaning? We all owe it to our students and ourselves to be crystal clear on our goals and expectations—no surprises and no excuses.



Clear Targets. Learning targets for students not only need to be appropriate, they need to be clear. It's easy to agree on a target like "communicates well." But, what does this mean? What type of communication, in what contexts, for which purposes? The key to effective student learning targets, be they at the national, state, district, or classroom level, is that they are specific enough to enable everyone to share the same understanding of what students need to know and be able to do. When targets are ambiguous, instruction can take students to vastly different places and assessments can be vastly different. The goal here is *not* to standardize instruction; rather, the goal is to aim at the same learning target even if teachers have different instructional designs.

Learning targets also need to be clear enough so that the persons who find or write assessment items and tasks have the same interpretation of what should be covered as the persons who wrote the target statements. Will future teachers interpret them the same? Are the interpretations clear enough so that, when the assessment results are used to profile achievement strengths and weaknesses, users will know what to do about it? For example, how would one design an assessment for "communicates well?" The assessment could be anything from writing an essay to observing students on the playground as they informally communicate with their peers.

Targets can be unclear for lots of different reasons. In our experience when trying to design assessments to match ambiguous content standards, we have noted the following sources of confusion:

1. Is it clear what cognitive level is being assessed? Is it recall of facts? Higher-order thinking? For example, the word "know" in targets can be interpreted in many ways. How will students act when they "know?" Rote recall of facts, like restating definitions? Or is "understanding" demonstrated through picking out new examples (or counter examples) of the principle, restating the principle in one's own words, or independently using the principle when engaged in real-life tasks?

Caution...striking a balance between detail and over-restrictiveness can be tricky. We want targets clear enough that we can all agree on what student success looks like, but not so detailed that "these 10 things" are all we mean. Or, even worse, that "these 300 things" are exactly what we mean.

2. How will targets fit together across grades? For example, "Writes for various audiences and purposes." How will this be different in grade 3 than grade 8?

3. How will assessment of targets repeated across subject areas be handled? Common repeated targets are group skills, oral presentations, critical thinking, study skills, interpreting and using graphs, etc. Is there a fundamental difference in how these repeated targets are expressed in different subjects? If so, how will the assessments be different? If not, how will the assessments be the same? How will instruction and assessment complement each other to give a full picture of all aspects of repeated targets? Should repeated targets be assessed as part of subject area assessments, or separately?

4. Are there any holes in the statement of essential learning targets for students? Are there important targets not covered anywhere? Do targets reflect current thinking about what expertise looks like? A standards-based assessment can be only as good as the target statements on which it is based.

Rule of thumb: *Is the target clear enough that a group of teachers would agree on the range of knowledge, skills and performance implied by the target? Would they agree on what to teach and what to assess?*

Tricky? You bet. In fact, Joan Herman and her colleagues state that available evidence suggests that many states' standards currently are not strong enough to support rigorous assessment development. *But, we believe that it's the attempt to clarify targets, as much as having final clear targets in place, that makes a difference.* In groups we've worked with, the general consensus is that everyone who makes the effort ends up with a much more indepth understanding of what they are trying to accomplish with students.

Types of Learning Targets. There are a million (well, actually maybe a hundred) different ways to categorize the types of learning targets (achievement goals, outcomes) we've seen for students. But, hold on, you're saying, why would one even want to "categorize" them? Well, this isn't just an outgrowth of compulsiveness on the part of number-crunchers. The process of categorizing helps to do three things. First, it helps folks to thoroughly think through what they want students to know and be able to do (in other words, clarify targets). Second, it helps folks determine if they have a good 'mix' of learning targets. And, finally, it will help folks, later, to choose the appropriate assessment method.

Here are two different "takes" on how to categorize learning targets for students. The first thing to remember about these (or any other) categorization schemes is that they are conveniences made up by someone in order to help people discuss things that are complex. There is no "truth" out there in the universe that "there are five kinds of student outcomes" or that "there are two basic kinds of student outcomes, each of which has two variations." Each scheme has its strengths, weaknesses and interesting aspects. Neither is perfect.

Take One: Bob Marzano. Marzano divides student outcomes into two types: process skills and content/declarative knowledge skills. Both process skills and content knowledge can be simple or complex. Simple process skills are short routines that are applied fairly consistently across situations, such as long division. Complex process skills are those that require many decisions and the integration of many simple process skills, such as writing or critical thinking. Likewise, simple content knowledge relates to things like recall of facts, while complex content knowledge relates to understanding concepts and making generalizations. **Figure 1.2** shows examples of each of these types of student outcomes.

Reference Box:

For more information on Bob Marzano's categorization scheme see: Robert J. Marzano, 1996, "Understanding the Complexities of Setting Performance Standards," in Robert Blum and Judy Arter (Eds.), *Student Performance Assessment in an Era of Restructuring*, Section I, Article #6. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723. ISBN 0-87120-267-0.

Figure 1.2
Types of Student Learning Targets

Type of Learning Target	Examples
Process Skills: Simple	Long division Punctuation, grammar Decoding words
Process Skills: Complex	Problem solving Writing Setting up an experiment Critical thinking Group cooperation Lifelong learning Dance
Content/Declarative Knowledge: Simple	Recall facts—e.g., dates, places, events
Content/Declarative Knowledge: Complex	Concepts—e.g., democracy Generalizations—e.g., "power corrupts"

Take Two: Rick Stiggins. Rick Stiggins finds that classifying student learning targets into five categories helps teachers find a good mix in instruction and assessment:

- **Knowledge Mastery**—Knowing and understanding substantive subject matter content, including facts (e.g. "John Kennedy was assassinated on November 22, 1963"), generalizations (e.g. "People holding high political office put their lives in jeopardy"), and concepts (e.g. "political assassinations").

- **Reasoning Proficiency**—The ability to use content understanding to reason and solve problems. Reasoning includes things such as analyzing, comparing, thinking critically, and decision making.

🕮 **Reference Box:**

For further reading on types of targets à la Rick Stiggins, see: Rick Stiggins, 1997, *Student Centered Classroom Assessment*, Chapter 3. Prentice-Hall, (201) 236-7000.

119

- **Skills**—Doing things such as reading fluently, working productively in a group, making an oral presentation, speaking a foreign language, or designing an experiment.
- **Ability to Create Products**—Creating tangible products such as an essay, a research reports, visual art, or a wood table.
- **Dispositions**—Student attitudes, including: attitude toward school, civic responsibility, self-confidence, desire to learn, flexibility, and willingness to cooperate.

Caution...

Affective targets can be a red flag in some communities. If so, the user can delete any references to the affective domain in this chapter and stick just to the cognitive domain.

Related *Toolkit98* Chapters and Activities:

Activity 1.2—*Clear Targets—What Types Are These?* Asking people to describe what kind of target they are looking at (knowledge, reasoning, etc.) is an excellent way to begin to tease out differences in what is meant by targets statements.

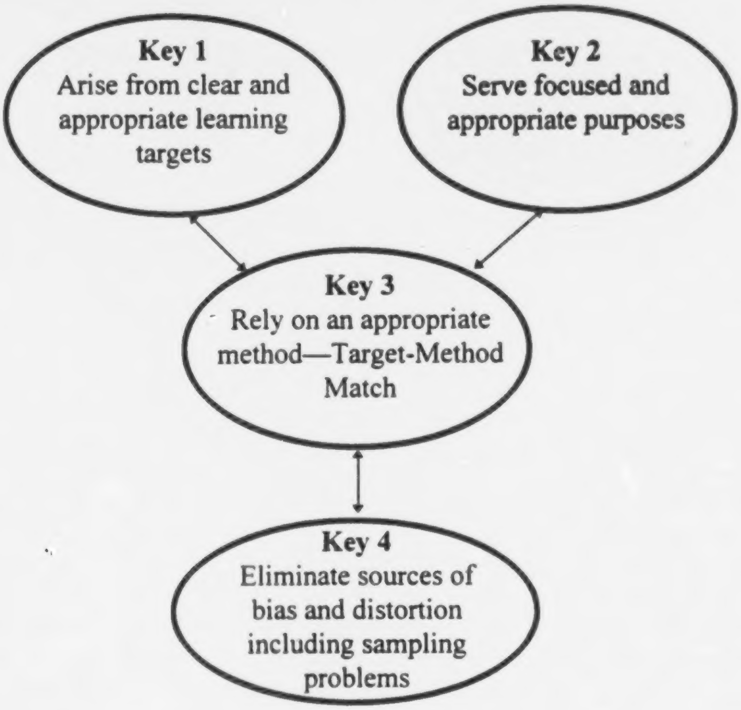
Activity 1.6—*A Comparison of Multiple-Choice and Alternative Assessment.* Comparing different ways to assess a content area provides a means of exploring what it means to know and understand a content area.

Activity 1.10—*Clear Targets and Appropriate Method—The View From the Classroom.* In this activity, participants are asked to self-evaluate the clarity of their learning targets using a rating form.

Activity 2.1—*Sorting Student Work.* Analyzing what makes student work effective is an excellent way of opening up the discussion of what it means, for example, for a student to write well.

BEST COPY AVAILABLE

Key 2: Focused and Appropriate Purpose


Key 1
Arise from clear and appropriate learning targets
Key 2
Serve focused and appropriate purposes
Key 3
Rely on an appropriate method—Target-Method Match
Key 4
Eliminate sources of bias and distortion including sampling problems

We'll talk about two things in this section. First, we'll address *why* educators assess—purposes. In other words, who are the users and uses of assessment results? This will underscore the importance of doing a good job of assessing students. Then, we'll look at "the rest of the story"—how purpose actually affects the way an assessment is designed.

So, why do folks assess student achievement? Who uses the results and what do they use the results for? Well, just about everybody for just about everything, and assessment activity seems to get more intense every day. For example:

- **Teachers** assess students day-to-day in the classroom for such purposes as: planning instruction, evaluating what worked and what didn't, grading, promoting student self-control of progress, and communicating with parents.
- **Students** use the results of assessments to decide what they'll study, how much they'll study, whether it's even worth studying, what they're good at (or not), their self-worth, who they might associate with, and how they might earn their living as an adult. Assessment information can also give feedback to students so they can reflect on their performance and make changes. Ultimately, we want students to understand how they

learn, evaluate their own performance, and undertake the learning necessary to improve their performance.

- **Parents** use the results of assessments to determine rewards and punishments for their children, apportion family resources (who will go to college?), make judgments of family self-worth, decide whether to vote for a levy, and pick a place to live.
- **Principals and other administrators** use assessment results to promote and graduate students, allocate resources, plan professional development, and report to the public how well the school or district is doing.
- **State departments of education** use assessment results to report to the public, levy rewards and sanctions for districts, and distribute resources.

Looking at this list reminds one that these are pretty important uses for assessment results and that we all had better be darned sure that our assessments are of good quality. What would happen, for example, if an assessment gave an inaccurate picture of student achievement? What would happen if what was *actually* assessed was not really what was *thought* to be assessed? Or worse yet, what if users were unsure as to what they were assessing, so they didn't know what the results really meant? Is everyone *positive* that they are accurately assessing the most enduring outcomes for students so that the decisions made can really serve to guide learning?

Looking at the list of users and uses also reminds one of the crucial importance of not only large-scale assessments—those that occur in roughly the same way at roughly the same time across classrooms—but also *classroom* assessments. After all, which assessments—day-to-day classroom assessments or once-a-year large-scale assessment—most affect the kinds of decisions made by teachers, parents, and students? (We would choose classroom assessment; hopefully readers did too.) What happens if classroom assessments are not well thought-out and executed?

The point here is not to suggest that teachers should be blamed for lapses in their knowledge about classroom assessment. (After all, most teachers never had the opportunity to learn about assessment because most states don't even require an assessment class for certification. And, even in places where an assessment course is required, there is an evolving understanding of what teachers really need to know and be able to do to be good classroom assessors.)

The point *is* that student assessment is of crucial importance. That's why there is activity on all fronts to improve it, from clearly defining valued student learning targets at the state level and rethinking how to best assess them in large-scale assessment, to changing coursework for pre-service teachers, to assisting teachers to align important learning targets to instruction and assessment in the classroom. *One of the efforts to assist teachers to fine tune classroom assessment practice is this Toolkit98!*

Now for the second part of this section—**How does purpose affect how educators assess?** It's probably obvious that an assessment good for one purpose—for example, providing detailed diagnosis of a student's strengths and weaknesses—is not necessarily best for other purposes—for example, determining the strengths and weaknesses of the school's overall curriculum or whether most students attained the school's grade-level goals for student performance. Thus, a first major decision to make is deciding one's purpose for assessment.

Just consider the differences in uses of information from large-scale and classroom assessments. In general, large-scale purposes require more rigorous evidence of technical quality than do classroom assessments, primarily because important decisions are likely to be based on them and because it is usually only a single testing episode. In contrast, for classroom purposes, a teacher has lots of formal and informal evidence upon which to base decisions, and so the results of any single, faulty assessment are not likely to be given undue weight.

As other examples of how purpose can affect assessment design, consider these:

1. A single multiple-choice or short answer multiplication test may be perfectly acceptable to determine whether or not third graders have learned their multiplication facts, but would not be appropriate for making a decision about the overall quality of the third grade mathematics program.

2. A short answer or multiple-choice assessment designed to measure student knowledge of specific scientific principles might be useful for partially determining a student's grade. Inferring that this assessment sufficiently measures the student's ability to perform scientific tasks that call for an understanding of these principles would require the assessors to observe the student applying that scientific knowledge in a laboratory setting.

Related *Toolkit98* Chapters and Activities:

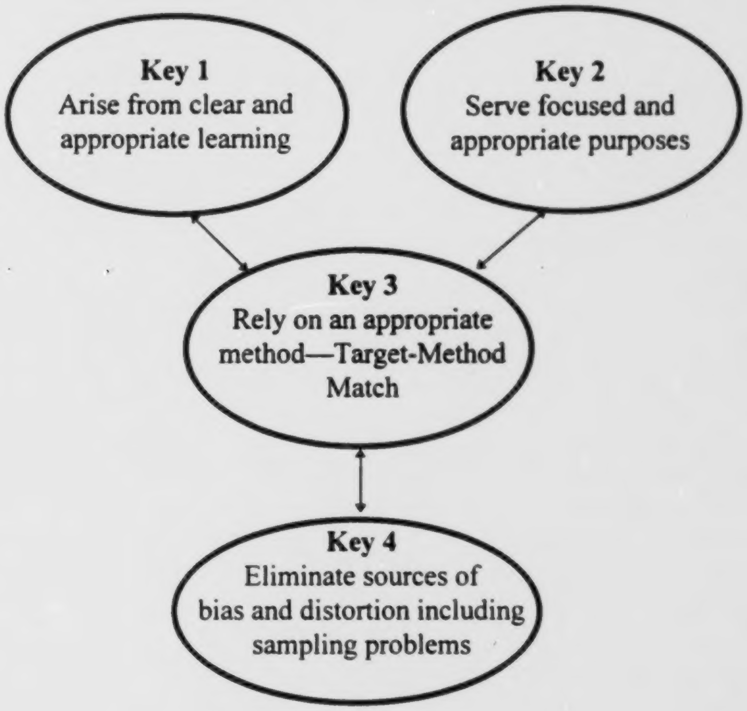
Activity 1.8—*Sam's Story* asks participants to judge the assessments they'd trust to give good information for a particular purpose—determining proficiency in math for instructional planning. At the end, it asks whether other purposes (for example, whether a student is working up to potential), might require different assessments.

Activity 1.9—*Going to School* asks participants to think about different designs for performance criteria and which might be most useful for different purposes.

Chapter 2 addresses various purposes for classroom assessment and posits design implications.

3. Some assessments are used mostly to gather information about students in order to make decisions about them—for example, grading or certifying competence. Other assessments are designed more to involve students in their own assessment and thus serve an instructional function. This distinction between assessment to "monitor" and assessment to "teach" has implications for assessment design. For example, rubrics used by teachers to monitor student performance might not need to be as detailed as rubrics used by students to learn and practice the features of writing that make it work.

Key 3: Appropriate Methods (Target-Method Match)


Key 1
Arise from clear and appropriate learning
Key 2
Serve focused and appropriate purposes
Key 3
Rely on an appropriate method—Target-Method Match
Key 4
Eliminate sources of bias and distortion including sampling problems

The third Key to Quality with respect to assessment is to match targets and purposes to methods. We are *not* of the opinion that the only good assessment is a performance assessment. Rather, there are times and places for all different forms of assessment. We'll again take two looks at "target-method-match," one from Bob Marzano and one from Rick Stiggins.

Take One: Bob Marzano. **Figure 1.3** shows Bob Marzano's scheme for matching targets to methods.

Figure 1.3
When to Use Alternative Assessment

Type of Skills	Multiple-Choice, Short Response	Performance Assessment
Process Skills: Simple	Long division Punctuation, grammar Decoding words	
Process Skills: Complex		Problem solving Writing Setting up an experiment Critical thinking Group cooperation Lifelong learning Dance
Content/Declarative Knowledge: Simple	Recall facts—e.g., dates, places, events	
Content/Declarative Knowledge: Complex		Concepts—e.g., democracy Generalizations—e.g., "power corrupts"

The basic premise is "simple target, simple assessment; complex target, complex assessment." Multiple-choice, matching, true/false and short answer are perfectly fine to assess simple processes and simple declarative knowledge. To assess **complex procedural skills**, Marzano would set up a task that requires students to use the skills in question and develop performance criteria to measure different levels of successful performance.

To assess **complex declarative knowledge** Bob Marzano recommends:

1. If one wants to see how well students understand a concept, say "democracy," one has to watch them use their knowledge in some way. So, one needs to decide how students will apply the knowledge—will they compare, classify, evaluate cases, deduce consequences, construct support, or do something else? For example, if the goal is to assess student understanding of the concept of "democracy" one could ask students to analyze cases—for example, they might identify current instances of countries claiming to be democracies and analyze the extent to which each actually *is* a democracy.

2. Then one needs to decide how students will report the results once they apply the knowledge in question—will they write, speak, draw, or something else? For the democracy example students might write a paper detailing their analysis.

3. Finally, one must develop performance criteria for three things:

    a. Content/declarative knowledge—how well does the student know the content? For the democracy example, what would be indicators of understanding the concept of "democracy": a sound definition; selection of appropriate countries; attention to the most telling details?

    b. Quality of the product—how well did the student present the work in writing, speaking, etc. For the democracy example, what is the quality of the paper on democracy in terms of organization, appropriate voice, word choice, sentence fluency and conventions? Did these support the purpose (arguing a case) and the major thesis of the paper?

    c. Quality of the application—how well did the student execute the knowledge application process? For our running example, how good was the student analysis and comparison? What makes a good analysis or comparison? What are the indicators that good analysis or comparison occurred?

Take Two: Rick Stiggins. Now, for "target-method match" according to Rick Stiggins. He maintains that, although you can assess most types of student learning targets by most methods, there are some more and less efficient ways to do it. For example, if all you want to know is whether students know their multiplication facts, why design performance assessments? **Figure 1.4** shows his recommendations for matching targets to methods. X's denote a good match; O's denote a partial match.

Figure 1.4
Aligning Achievement Targets and Assessment Methods

	Selected Response	Essay	Performance Assessment	Personal/Oral Communication
Knowledge Mastery	X	X		O
Reasoning Proficiency	O	X	X	X
Skills			X	X
Products		O	X	X
Dispositions	X	O	O	X

It's simply not efficient to use performance assessment or personal oral communication to assess every *knowledge* outcome educators have for students. For example, using performance assessments to see whether students know all their multiplication facts could take years. But we *could* assess instances of ability to multiply in the context of a problem solving performance task.

While it is possible to assess some kinds of student *reasoning* skills in, say, multiple-choice format, to really see reasoning in action one needs a more complex assessment format. For example, most standardized, norm-referenced tests have questions about such things as fact versus option and "what is most likely to happen next." But these are usually assessed out of context as a discrete skill. One would need a performance assessment to see how students can use all their reasoning skills together to address an issue, or to see if they know when, for example, they need to identify an opinion.

Knowledge *about* what it takes to *perform skillfully* or to *produce a product* can be assessed in multiple-choice format, but to actually *see* if a student can *do* it, one needs a performance assessment. (For example, as we discussed earlier in this chapter, one can assess student knowledge about how to give a good oral presentation through an essay, but if one wants to see if students can apply this knowledge, one has to have students *give* an oral presentation.)

Selected response questionnaires can tap student *dispositions*, but so can open-ended questions (essays) and personal communication with students.

Related *Toolkit98* Chapters and Activities:

Activity 1.6—*A Comparison of Multiple-Choice and Alternative Assessment* provides participants an opportunity to compare a traditional multiple-choice assessment with a performance assessment and discuss when each should be used.

Activity 1.7—*Target-Method Match* introduces assessment methods and gives participants practice in matching methods to learning targets.

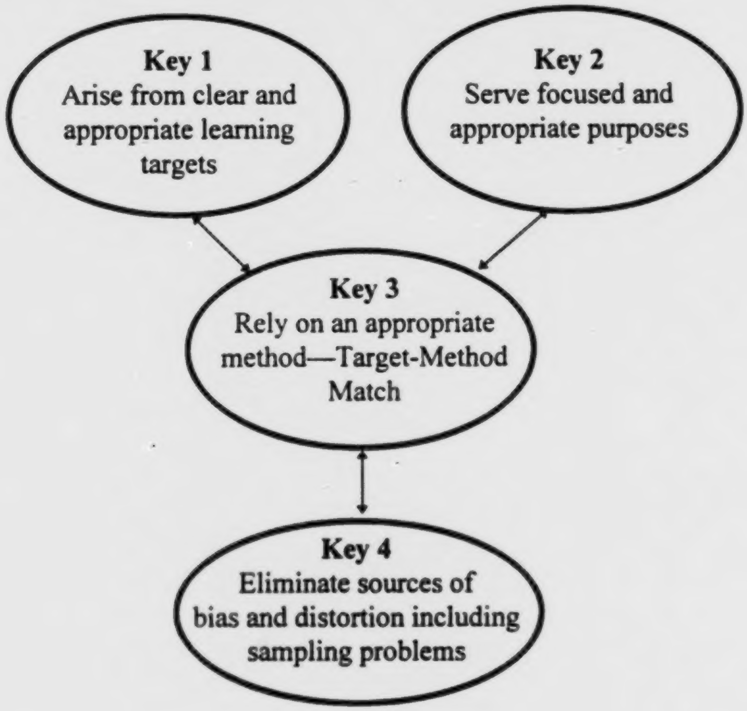
Activity 1.10—*Clear Targets and Appropriate Method—The View From the Classroom* asks participants to self-evaluate the extent to which they successfully match assessment methods to targets.

Chapter 3 covers design options for alternative assessment and includes additional discussions of target-method match.

Activity 3.2—*Spectrum of Assessment Activity* looks at how to "open up" traditional assessment tasks in order to measure more complex outcomes.

Activity 3.4—*Assessing Learning: The Student's Toolbox* demonstrates the relationship between assessment tasks and the student learning trying to be assessed.

Key 4: Eliminating Sources of Bias & Distortion


Key 1
Arise from clear and appropriate learning targets
Key 2
Serve focused and appropriate purposes
Key 3
Rely on an appropriate method—Target-Method Match
Key 4
Eliminate sources of bias and distortion including sampling problems

Assume that targets are perfectly clear and appropriate, purposes have universal agreement, and the absolutely best way to assess each target has been picked. Is this all? Unfortunately, no. It's still possible, even easy, to execute the plan poorly.

In short, things can go wrong in assessment. Have you ever tried to engage students in an instructional activity and not gotten at all what you expected to? The instructions weren't clear, or there wasn't enough time, or students didn't have all the prerequisite skills, or the activity didn't allow students with different learning styles to do their best.

Well, the same thing can happen in assessment. These "things that go wrong" are called sources of mismeasurement, bias and distortion, or invalidity. The result is that the information from the assessment doesn't mean what we think it means. What happens if the ability to read the instructions interferes with a student's ability to demonstrate math skills? Or, the necessity to write a response interferes with how well a student can demonstrate ability to set up a scientific experiment?

Then these assessments could really measure reading or writing rather than math or science—these are serious potential sources of bias and distortion. Or, what happens if student writing competence is judged from only a single piece of writing? Then the judgment of student writing ability may be biased because not enough *samples* of student writing were obtained across audiences, purposes, and content to really determine, in general, writing competence.

If these assessments then form the basis for a grade or for certifying competence on a state graduation test, the result would be unfortunate. A grade or certification of competence is only as good as the assessment upon which it is based.

This, then, is the fourth Key to Quality—attending to what might go wrong and fixing it. Since this topic is handled so completely elsewhere in *Toolkit98*, we'll not cover that ground again here.

⊂ Related *Toolkit98* Chapters and Activities:

Eliminating potential sources of bias and distortion is discussed in many ways in many places in *Toolkit98*. **Chapter 3** contains an in-depth discussion of various alternative assessment designs and the relative merits of each approach. Much of this discussion relates to issues of potential sources of bias and distortion.

Activities that stress the importance of quality include: **Activities 1.5**—*Clapping Hands* (potential sources of bias and distortion in performance assessments); **1.8**—*Sam's Story* (the most valid pieces of information for a particular purpose); **1.11**—*Assessment Standards;* **1.12**—*Assessment Principles* (equity); **3.1**—*Performance Tasks, Keys to Success* (characteristics of quality tasks); **3.3**—*Performance Criteria, Keys to Success* (characteristics of quality criteria); **3.6**—*How to Critique an Assessment* (practice critiquing on all aspects of quality); **3.7**—*Chickens and Pigs* (equity); **3.8**—*Questions About Culture and Assessment* (equity); and **3.9**—*Tagalog Math Problem* (equity).

Summary of Steps to Quality

We've had a brief excursion into keys that describe quality assessment in any context, for any purpose, for all types of assessment. The remainder of *Toolkit98* focuses on alternative assessment (particularly performance assessment), not because we don't think other forms of assessment are worthwhile, but because that's the area in which most people want assistance right now. So, we'll summarize our look at the Keys in terms of when each occurs in the alternative assessment design process. Many of these steps are covered in more detail throughout *Toolkit98*. We've noted where in *Toolkit98* each is expanded upon. Some topics are beyond the scope of *Toolkit98*. (We have noted those that are not addressed in detail.)

1. Clearly define what it is you want to assess. (This is **Key 1**—*Clear and Appropriate Achievement Targets.*) In *Toolkit98* we focus primarily on those targets most appropriately assessed by alternative assessment procedures.

2. Clearly state the purpose for the assessment, and don't expect one assessment to work for other purposes for which it was not designed. (This is **Key 2**—*Clear and Appropriate Purposes.*) In *Toolkit98* we focus primarily on classroom purposes for assessment such as monitoring student achievement, planning instruction and engaging

students in self-assessment. These purposes are expanded on in **Chapter 2**—*Integrating Assessment with Instruction.*

3. Match assessment methods to the achievement target(s) and purpose. (This is **Key 3**—*Appropriate Methods.*) Developing alternative assessments for classroom assessment purposes is the topic of **Chapters 2**—*Integrating Assessment with Instruction* and **3**—*Designing High Quality Assessments.*

4. Specify illustrative tasks (activities, problems) that would require students to demonstrate the desired skills and accomplishments. Avoid tasks that may only be interesting activities for students, but may not yield evidence of a student's mastery of the desired outcomes. (These are **Key 3**—*Appropriate Methods* and **Key 4**—*Avoiding Potential Sources of Bias and Distortion.*) Design options and quality considerations for tasks is in **Chapter 3**—*Designing High Quality Assessments.*

5. Specify the criteria and standards of judging student performance on the task selected in step 4. Be as specific as possible, and provide samples of student work that exemplify each of the standards. Develop a reliable rating process that would allow different raters at different points in time to obtain the same, or nearly the same, result. If used in the classroom by a single teacher, the rating system must allow consistency across students.

   (These are **Keys 3**—*Appropriate Methods* and **4**—*Avoiding Sources of Bias and Distortion.*) Criteria and development of scoring processes are dealt with extensively in **Chapters 2**—*Integrating Assessment with Instruction* and **3**—*Designing High Quality Assessments.*

6. Avoid other pitfalls that can lead to mismeasurement of students. (This is **Key 4**—*Avoid Potential Sources of Bias and Distortion.*) A gentle introduction to technical issues, especially as related to classroom assessment, is found in **Activity 1.5**—*Clapping Hands.* Equity and other quality issues are explored in **Chapter 3**—*Designing High Quality Assessments.*

7. Collect evidence/data that shows that the assessment is reliable (yields consistent results) and valid (yields useful data for the decisions being made). For performance assessments, this might be demonstrated through the level of agreement between scores given by different assessors for the same student work, and evidence that students who perform well on the assessment also perform well on other related items or tasks.

   This topic is not covered in detail in *Toolkit98.*

9. Ensure "consequential validity"—the assessment maximizes positive side effects and minimizes negative ones. For example, the assessment should give teachers and students the right messages about what is important to learn and to teach; it does not

restrict the curriculum; it is a useful instructional tool; and the decisions made on the basis of the assessment results are appropriate.

Consequential validity and the messages our assessments send teachers, students, and parents are discussed extensively in **Chapters 2**—*Integrating Assessment with Instruction* and **3**—*Designing High Quality Assessments.*

10. Use test results to refine assessment and improve curriculum and instruction; provide feedback to students, parents, and the community. **Chapters 2** and **3** further discuss the relationship between assessment and instruction; **Chapter 4**—*Grading and Reporting, A Closer Look* addresses grading and reporting issues.

Chapter Summary and Conclusion

This chapter provides the reader with an overview of the reasons changes are being made in assessment, and with a beginning understanding of the issues assessors must grapple with as they use (or develop) assessments to make educational decisions about students and programs. The remaining chapters provide a more indepth discussion of these topics.

Assessment is changing because education is changing. Research shows that what students need to know and to be able to do, and the way in which knowledge and skills need to be taught, are often quite different from conditions in the past. This is not an indictment of schools. Changes in the world are happening at such a rapid pace that few aspects of society have been able to keep up. But keep up we all must, if our students are to be successful in the changing world of tomorrow.

Although we present a strong case for alternative assessment in *Toolkit98,* we neither say that all assessments need to be of this type nor reject the use of multiple-choice and other forms of selected-response tests. We do affirm that alternative assessments, when designed well, offer appealing ways to assess complex thinking and problem-solving skills and, because they are grounded in realistic problems, are potentially more motivating and reinforcing for students. However, while alternative assessments may tell us how well and deeply students can apply their knowledge, selected-response (e.g., multiple-choice tests) may be more efficient for determining how well students have acquired the basic facts and concepts. Educators need to choose assessments based on professional information, not on tradition ("that's how I was tested"), personal opinion ("I like them") or trendiness ("Grant Wiggins says they're great.") A balanced curriculum requires a balanced approach to assessment.

The need for high quality assessment information to make informed decisions about changes in students and programs will be critical to the success of the educational improvement effort. There is no single correct method for assessing students, but there are ways in which all forms of assessment can be used well to help schools make good decisions and meet students' needs. Keys to quality include: clear and appropriate learning targets; clear and appropriate purposes; matching method to purpose and target; and avoiding potential sources of bias and distortion.

Activity 1.1

Changing Assessment Practices: What Difference Does It Make for Students?

Purposes:

1. To stimulate thinking and discussion among teachers, parents, and diverse communities of learners about the practical, learning-centered reasons for changes in assessment practices
2. To prompt reflection about current practice as a foundation for action

Uses:

This is a good introductory activity for looking at the rationale for changes in assessment. It can be used to begin a PTA or other family and community session. It is also useful with a school faculty. It has been used in a large, auditorium-style session with several hundred participants.

Rationale:

As use of a variety of assessment tools expands, it's essential that educators look together at assessment practices. As educators rethink and improve the quality of assessment practices they need to share their growing understanding of the impact of assessment practices on children's learning, and, more fundamentally, on the belief of children in their ability to learn.

This activity is built around the work of a group of British Columbia educators who developed a guide to communicate with parents and community members about the impact of assessment practices on the learning of young children and the rationale for changes underway in primary classrooms.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A1.1,O1**—*Changing Assessment Practices, Purposes*, **A1.1,O2**—*Questions for Parents*, and **A1.1,O3**—*Questions for Educators*
- **Handout A1.1,H1**—*Assessment/Evaluation/Reporting*

Time Required:

20-30 minutes

Facilitator's Notes:

1. (1 minute) Set the stage for reflection and discussion by noting that participants will be examining and discussing a sheet that shows one group of educators' way of helping parents understand the specific changes in assessment practice that are taking place in classrooms and how directly those changes are related to children's learning and academic self-concept. The sheet shows what this group of educators has come to value and why, as well as the actions that flow from their understanding. Use **Overhead A1.1,O1**—*Changing Assessment Practices, Purposes* to describe the purposes of the activity, as needed.

2. (5 minutes) Provide copies of **Handout A1.1,H1**—*Assessment/Evaluation/Reporting* and ask participants to read it while thinking about the questions on **Overhead A1.1,O2**—*Questions for Parents* or **A1.1,O3**—*Questions for Educators*.

3. (10 minutes) Ask participants to find a partner and discuss the handout and their thoughts about the questions on the overhead. Are there some common questions? Common areas of agreement?

4. (5 minutes) If time permits, ask for volunteers to report key ideas, issues, and concerns from their discussions.

5. (5 minutes) Remind participants that this activity is intended *not* to present the specific practices that educators *must* use (and thus end conversations about our assessments), but to open or extend ongoing conversations about day-to-day assessment practices and student learning.

   You might close by asking participants to create their own list of reasons for change in assessment practices—or—write down one insight they gained from the discussion or one question they want to learn more about.

Changing Assessment Practices

Purposes:

1. To stimulate discussion about the practical, learning-centered reasons for changes in assessment practices

2. To prompt reflection about current practice as a foundation for action

Overhead A1.1,O1

Questions for Parents

1. Which changes are ones that leap out as ones you'd say, "Of course, it really needs to be this way to work best for our children?"

2. Are there any that cause you to pause to think about whether the change is one you understand and want to see, or would like to hear more about?

3. Do you have any experiences with the changes that you can share to help make them real for other parents? Have your children brought home things that illustrate any of these changes?

Overhead A1.1,O2

Questions for Educators

Read the rows across, noting the sets that stand out for you as you think about your own or your school's classroom assessment practices. Circle the practices in the left column that are in use now in your school or classroom. Put a star next to the practices in the third column that you are moving toward or using now.

1. Which of the assessment practices in the third column are ones that leap out as ones about which you'd say, "Of course, it really needs to be this way to give our students the best opportunity to show what they know and can do?"
2. Are there other practices that cause you to pause and think about whether the practice is one you understand and want to see, or would like to learn more about?
3. Are there practices in the first column that need to be re-examined? Eliminated in your setting?
4. What experiences can you share to help make the assessment practices real for others? What are you doing in your work that illustrates specific changes?

Overhead A1.1,O3

Assessment/Evaluation/Reporting[1]

We used to...	***But...***	***So now...***	***Because...***
place more emphasis on what students could not or should not do	we learned that this focus undermined the confidence of many students and that we could be more supportive of their accomplishments	we begin with what students can do, then consider their learning needs	this helps them to develop confidence and gives a foundation for building and further refining skills and knowledge
fail students who did not meet pre-set expectations for behavior or ability to do tasks	we found that some students doubted their ability to learn and this increased the probability of their dropping out of school	teachers give students the support needed to allow them to make continuous progress	this maintains their self-esteem and confidence, thus prompting further learning by strengthening the disposition to learn
use pencil/paper tasks as the main way of assessing and evaluating students	we now know that this gave a limited view of what students could do	we encourage students to represent their learning in a variety of ways (show what they know)	this provides opportunities for more students to demonstrate their intelligence and to be successful learners
compare learners to each other	this made comparisons more important than the actual learning	each learner is evaluated on what he or she can do in relation to expectations and skills that are widely held as being important, continually refined, and purposefully applied	this helps each student feel valued as a learner and builds on **individual** strengths, which encourages a good start toward lifelong learning
use checklists for students' report cards	they gave limited information about what students could do	we use information from observations, conferences and collections of students' work to develop anecdotal reports	they give more comprehensive information about what students can do
use letter grades for reporting students' progress (A, B, C) (G,S,NI)	letter grades were dependent on teacher and parent interpretation and often focused on surface knowledge rather than understanding	we use anecdotal reports to describe students' learning	they give a more detailed picture of what students can do and identify future learning goals
exclude students from the assessment and evaluation process	this did not encourage the development of self-evaluation skills	students are encouraged to take a more active role in assessing and evaluating their own progress and, with the help of the teacher, set future learning goals	as students construct meaning of the world around them, this process encourages self-evaluation, independent learning and a commitment to further learning
plan conferences for parents and teachers to exchange information	this often overlooked the people with the most relevant information—the students as developing learners	teachers are beginning to plan ways to include students in the conference with parents	together, they can develop a shared understanding of students' abilities, interest and learning needs, resulting in the setting of realistic learning goals

Handout A1.1,H1

---

[1] From *Supporting Learning: Understanding and Assessing the Progress of Children in the Primary Program*. Province of British Columbia, Ministry of Education, 1993. Permission granted for use in training.

BEST COPY AVAILABLE

Activity 1.2

Clear Targets—What Types Are These?

Purposes:

1. To consider different kinds of learning targets for students
2. To consider what a good mix of learning targets for students might be
3. To become familiar with local content standards
4. To experience how frequently learning targets for students that we thought were clear really aren't

Uses:

This activity is at the intermediate level—best used after reading the **Chapter 1** text and doing other context and stage setting activities in this chapter, such as **Activity 1.1**—*Changing Assessment Practices, What Difference Does It Make For Students?* or **1.12**—*Assessment Principles*. It develops skill related to **Chapter 1, Key 1**—*Clear Targets*.

Rationale:

This activity makes several points, so there are several rationales for it.

Rationale #1: State and local content standards are intended to clarify targets so that instruction across the system is focused on what is important. We've found that it is most meaningful to teachers to have them list their goals first so that they can see that the state or district content standards represent what they already value for their students. This activity relies on this method.

Rationale #2: We are coming to the conclusion more and more that **the single biggest barrier to improving classroom assessment** has nothing to do with the physical aspects of assessment at all (such as selecting the correct assessment method, making sure that there is only one correct answer, etc.). The biggest **barrier is not having clear learning targets for students**. One can't assess (or teach) something if one doesn't know with crystal clarity what it is they are trying to assess (or teach).

For example, every set of content standards we've seen includes "critical thinking." Well, what does that mean? What does a person look like when they are thinking critically? What does such a person know and what is such a person able to do? Likewise, what is "communication in math," "quality writing," or "life-long learning?" How do educators know when students are doing them well? How do educators know when students are developing adequately toward proficiency? This activity is only the opening gambit on this issue. Other activities in *Toolkit98* are also designed to assist teachers to be masters of the targets of instruction embodied in content standards.

Rationale #3: One reason that targets tend to be unclear is that, in the past, many teachers have had an "activities" focus to instruction rather than an "outcomes" focus. What we're trying to do here is to assist teachers to change their focus from "task" to "targets." This is not meant to disparage teachers, the importance of change in focus—to a standards-based approach to instruction—is only now becoming apparent.

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens
- **Overheads A1.2,O1**—*Clear Targets Purposes* and **A1.2,O2**—*Kinds of Achievement Targets*
- **Handouts A1.2,H1**—*Kinds of Achievement Targets*; **A1.2,H2**—*Examples of Learning Targets*, and **A1.2,H3**—*What Types of Student Learning Targets Are These?*

Time Required:

40-60 minutes

Facilitator's Notes:

1. (2 minutes) Use **Overhead A1.2,O1**—*Purposes* to introduce participants to the activity. The first step in designing any assessment is to have clear and appropriate learning targets for students. Educators can't assess something if they don't know what it is. Further, lots of learning targets for students are unclear—teachers disagree on what they mean. This activity provides an opportunity for participants to experience what "unclarity" is like.
2. (5-10 minutes) Ask participants to individually list their learning goals for students. (Use whatever language they are most familiar with—student learning outcomes, goals, content standards, or, merely, "What would you like students to know and be able to do when they leave your class?") Then ask participants to prioritize the top 5.
3. (5 minutes) Make a list of outcomes on a blank transparency or chart pack paper. An efficient way to do this is to randomly select 15 people and ask them to read the top item on their list. Then ask if anyone has any other items that *must* be added.

4. (5 minutes) Show **Overhead A1.2,O2**—*Kinds of Achievement Targets* and **Handout A1.2,H1**—*Kinds of Achievement Targets*. Describe the various types and show how the list generated by participants corresponds to the categories.

5. (15 minutes) Pass out a part of the state/local content standards (or course achievement targets) on which you'd like to focus. Ask participants to classify several targets. **Or**, use **Handout A1.2,H2**—*What Types of Student Learning Targets Are These?*

   Participants will often begin to "quibble" as to the "correct" classification. This occurs because people interpret the outcome differently; in other words, it is not clear what the standard means. This actually brings home the message of the entire segment—if educators don't agree on what targets mean, how can they exert a concerted effort to assist students to reach them? And, if educators don't agree on what targets mean, assessments will look quite different.

   In *every* situation where we've asked participants to look at local content standards or course outcomes and explain what they mean, there have been differences of opinion. Adopted standards (as well as everyday course outcomes) are likely to contain ambiguities. Therefore, district administrators and teachers have to come to consensus on their meaning—to openly discuss various interpretations to come to agreement on their breath and depth. Rather than being an occasion for despair, many educators have told us that such discussions are invaluable for influencing instruction, helping teachers feel more in control, and developing the "right" assessment. (Note: If the targets being clarified are *official district or state standards*, it is the responsibility of the district or state administration to clarify them and provide guidance so that a consensus is reached on what is intended. An individual group of teachers should not have to take on this responsibility unless requested and empowered to do so.)

   Caution

   Affective targets may be a red flag to some community members. If so, the entire activity can be done merely with cognitive targets. Simply remove all mention of the affective domain on all handouts and overheads.

6. (15 minutes, optional) If participants are looking at their local or course outcomes, ask them to decide if they have a good mix of types. What would a good mix look like?

7. (5 minutes, optional) Show an analysis of the learner targets in a textbook (**Handout A1.2,H3**—*Examples of Learning Targets*). Searching through this particular textbook resulted in only two reasoning targets, and these were in the "expanded" activities. The point: it was hard to find reasoning and product targets, so if teachers are counting the textbook to match our important learning outcomes, it could be misplaced trust.

Clear Targets

Purposes:

1. To consider different kinds of learning targets for students
2. To consider what a good mix of learning targets for students might be
3. To become familiar with local content standards
4. To experience how frequently learning targets for students that we thought were clear really aren't

Overhead A1.2,O1

Kinds of Achievement Targets

- Master Factual and Procedural *Knowledge*
- Use Knowledge to *Reason* and Solve Problems
- Demonstrate Mastery of Specific *Skills*
- Create Quality *Products*
- Acquire Positive *Affect/ Dispositions*

Adapted from Rick Stiggins, Assessment Training Institute, Portland, OR, 1995

Overhead A1.2,O2

Kinds of Achievement Targets

Master Factual and Procedural *Knowledge*

Some to be learned outright
Some to be retrieved using reference materials

Use Knowledge to *Reason* and Solve Problems

Critical Thinking
Problem Solving
Decision Making
Analytical, comparative, inferential, and/or evaluative reasoning

Demonstrate Mastery of Specific *Skills*

Speaking a second language
Giving an oral presentation
Working effectively on a team
Science process skills

Create Quality *Products*

Writing samples
Term projects
Artistic products
Research reports
Shop projects

Acquire Positive *Affect/Dispositions*

Positive self-concept
Desire to learn/read/think critically, ...
Positive attitude toward school
Good citizenship
Respect toward self and others
Flexibility
Perseverance

Adapted from Rick Stiggins, Assessment Training Institute, Portland, OR, 1995

Handout A1.2,H1

Examples of Learning Targets

Taken from the 4th grade Social Studies text *Discovering Washington.* (From Jan Chappuis, Central Kitsap School District, Silverdale, WA, 1996.)

Knowledge Targets (factual)

- Name at least two regions of which Washington is a part.
- Tell at least one way that location has affected Washington's history.
- Name five main parts of Washington State and describe key land and water forms of each.
- Define the term *natural resources* and give examples of those found in Washington.

Examples of key words: tell, understand, name, describe, list, identify, give examples.

Knowledge Targets (procedural)

- How to use map scales to measure distances.
- How to use latitude and longitude to locate places on a map or globe.

Examples of key words: how to (followed by some procedure).

Reasoning Targets

- Give examples of differences between coastal and plateau cultures and relate these to differences in the natural environment.
- Compare/contrast the pioneer way of life to their own.

Examples of key words:. Classify, compare, contrast, analyze, synthesize, determine, evaluate, justify, construct support, draw conclusions.

Skill Target

- Use map scales to measure distances.

Examples of key words: read, speak, assemble, operate, use, demonstrate

Product Target

- Make a relief map of any region of whole state, or make maps of products, points of interest, or land uses.

Examples of key words: create, design, make, produce, write, draw.

Handout A1.2,H2

What Types of Student Learning Targets Are These?

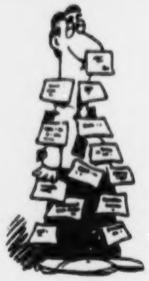
Instructions: Match the student learning targets on the right to Rick Stiggins' five types of learning outcomes on the left. There are no single correct answers, however, some responses are better than others. The important thing is to explain *why* you have categorized the target as you have. This type of matching will be important when we get to Step 2—matching target to assessment method.

Type of Learning Outcome	Student Learning Targets
a. Knowledge	1. Writes effectively for various audiences and purposes
b. Reasoning/Thinking	2. Develops knowledge and understanding of how language is used
c. Skills/Behaviors	3. Contributes productively to both large and small groups
d. Products	4. Uses historical and contemporary information to demonstrate cross-cultural understanding
e. Affect/Dispositions	5. Demonstrates knowledge of melody, rhythm, harmony, pitch, form, timbre, and tone quality
	6. Uses music vocabulary correctly
	7. Exhibits self-confidence, self-discipline, and self-motivation
	8. Develops a personal wellness plan that recognizes the impact of nutrition and fitness on personal well-being
	9. Chooses and devises appropriate strategies to solve mathematical problems
	10. Views self as capable of using mathematical skills and knowledge to make sense of new problem situations

Handout A1.2,H3

Activity 1.3

Post-it™ Notes



Purposes:

1. To expand or review understanding of assessment language
2. To develop a common language on which later activities can build

Uses:

Use in **Chapter 1** as a mixer and to review assessment terminology. It is a good introductory activity.

Rationale:

It's important to acknowledge the realities of the evolving vocabulary for assessment—and to point out the importance of crafting shared meaning in order to move forward.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead: A1.3,O1**—*Post-it™ Notes Purposes*
- **Handouts:** Small pieces of paper or Post-it™ notes containing definitions and terms (one possible set is included as **Handout A1.3,H1**—*Definitions/Terms*); the *Glossary* in Appendix E could be used as a handout
- Chart paper and marking pens (useful, but not necessary)

Time Required:

20 minutes

Facilitator's Notes:

1. (5 minutes) In this warm-up activity, participants examine assessment-related definitions and key terms. Each participant receives *either* a definition *or* a term. Their task is to find the partner who has the matching definition/term. (Slips of paper with definitions or terms can be made by cutting apart **Handout A1.3,H1.** If possible, one color paper could be used for definitions and another for terms.)

   The definition slips can be handed out individually as an opportunity for the facilitator to greet participants before the formal opening of the session, or can be included in individual packets of session materials.

2. (10 minutes) Once participants have connected with the individuals they believe are their partners, ask each pair to place their term and definition on a chart until all terms and definitions are on the chart. With a large group, you can divide into smaller groups and have a chart for each group. With a relatively small group, review the intended matches, and clarify, solicit challenges, and refine meaning. With larger groups, you can have a walk-about in which participants seek confirmation of their choices in the work of other groups.

3. (5 minutes) Close this brief activity with a summary of terms and, if appropriate, offer the opportunity for participants to write questions that still remain on a chart which will remain posted throughout the session. These questions often prompt significant thinking and give the facilitator some guidance about how to stretch following activities to take in the concerns of participants.

   The *Glossary* in **Appendix E** could be used as a handout to summarize these and other important definitions.

Post-it™ Notes

Purposes:

1. To expand or review understanding of assessment language
2. To develop a common language on which later activities can build

Overhead A1.3,O1

Definitions	Terms
An assessment that is given and scored in the same way for all.	Standardized Assessment
A question or task in which students must choose what they believe is the best or correct answer from a list of provided possibilities.	Selected-Response
A question or task in which students must create an answer or response themselves, but there is a specific correct answer or set of information that is expected.	Open-Response
Scoring performance as a whole; based on criteria, but not giving feedback on specific qualities of the student's work; giving a single score.	Holistic Scoring
A guide for scoring student performance; rules for scoring; sometimes includes descriptions of key characteristics of varying levels of performance.	Rubric

Definitions	Terms
Scoring performance on several different dimensions; giving more than one score.	Analytical Trait Scoring
A purposeful collection of student work to show effort, progress, or degree of proficiency.	Portfolio
A question or task that has many correct or excellent responses, intended to promote thinking and application of ideas and information. There is no single, specific correct answer.	Open-Ended
Statements of what students should know and be able to do.	Content Standards
Pre-established levels of proficiency to show mastery of content or skills.	Performance Standards
An assessment designed to reveal how an individual's performance compares to other similar students.	Norm-Referenced Assessment

Handout A1.3,H1, p. 2

Definitions	Terms
The key qualities or dimensions of an effective performance.	Performance Criteria
Short- or long-term activities, projects, or questions that serve as assessment opportunities.	Performance Tasks
An assessment designed to reveal what a student knows or can do in relation to a specific skill or body of knowledge.	Standards-Based Assessment
Assessment that is based on observation and judgment of student created products and/or performances; intended to provide a rich portrait of student learning.	Performance Assessment
Assessment tasks that resemble "real life" as closely as possible, are engaging for students, and reflect good teaching strategies.	Authentic Assessment

Handout A1.3,H1, p.3

Activity 1.4

Seeing Wholes*

Purpose:

To examine the connections among assessment, instruction and curriculum

Uses:

This activity is good as an opener (or brief reminder) in the early stages of a multi-activity professional development session to build a view of the rationale for changes in assessment, as described in **Chapter 1**.

It can be used with other activities (e.g., **Activity 1.3**—*Post-it™ Notes* and **Activity 1.6**—*A Comparison of Multiple Choice and Alternative Assessment)* to construct understanding about, and shared language for, alternative assessment.

Rationale:

Making clear connections between instruction and assessment is an essential part of the development and use of alternative assessments that will be used for ongoing communication with students and parents. In this activity, participants bring their own perspectives to a definition that could equally be attached to assessment, instruction **or** curriculum. This activity also tends to surface assessment issues.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A1.4,O1**—*Seeing Wholes Purpose*, and **A1.4,O2**—*A Round of Jeopardy*

---

* Developed by Dr. Joan Bycoff Baron at PACE, Harvard University, and used with permission.

Time Required:

10-20 minutes

Facilitator's Notes:

1. Use **Overhead A1.4,O1**—*Seeing Wholes Purpose*, as needed, to describe the goal of the activity.

2. (10 minutes) Show **Overhead A1.4,O2**—*A Round of Jeopardy* and have participants ask a question for which the statement is an answer. Write responses on an overhead.

   Here's an example of recently generated responses:

   - What is life?
   - What is a performance task?
   - What is curriculum?
   - What is ongoing assessment?
   - What is a system of coherent assessments?
   - What is a portfolio?
   - What is a good unit?
   - What is peer review?
   - What is education?
   - What is performance assessment?

3. (5-10 minutes) Points made by Dr. Baron to link the responses back to the concept of integrated curriculum and assessment:

   A good assessment system is the curriculum!

   A good set of assessment tasks has these features: they are coordinated, there is a sequence to them, kids learn the criteria, they become the owners of what quality work is, the act of learning the criteria is the curriculum. Learning events are the midwifery.

   Good curriculum is good assessment if criteria are specified.

Seeing Wholes*

Purpose:

To examine connections between assessment, instruction, and curriculum

---

*Developed by Dr. Joan Bycoff Baron at PACE, Harvard University, and used with permission.

Overhead A1.4,O1

A Round of Jeopardy

Answer:

A coordinated sequence of engaging and important learning events through which all children come to understand what high quality work looks like and how to produce it.

What is it?

Overhead A1.4,O2

Activity 1.5

Clapping Hands*

Purposes:

1. To develop guidelines for high quality performance assessments
2. To explore the importance of quality—the impact of our assessments on students

Uses:

This activity is at an early elementary level of difficulty. It can be used with **Chapters 1** or **3** to illustrate features of quality alternative assessment. It provides a gentle introduction to quality issues. It should be used prior to the more detailed quality activities in **Chapter 3—Activities 3.1, 3.3**, and **3.6**. In **Chapter 3** the activity can also be used to illustrate design options and the effects these can have on teachers and students.

Rationale:

Assessment is a deeply affective thing. It is about success and failure, about deeply held feelings, and about our sense of self-esteem. This activity is designed to show the impact of assessment both on performance and on the feelings at the receiving and rating ends of the process. It provides a graphic demonstration of the unintended consequences of how performance assessments are designed. It also results in a set of guidelines—do's and don'ts—for performance assessment.

We have successfully used this activity with teachers in all grades levels (K-12), and with both community college instructors and college of education faculty.

Materials:

- Overhead projector, screen, blank transparencies and pens; or chart paper; or chalkboard

---

[*] With thanks to Ruth Sutton for a demonstration of the activity and approval to use it. Contact Ms. Sutton at 29 Nevile Court, Nevile Road, Salford, M7 OPS, England, UK. And, with thanks to Jay McTighe and other members of the Maryland Assessment Consortium for ideas on involving the whole group in the activity.

- Ten volunteers from the group (7 if the group is small)
- Eleven chairs arranged in semi-circle—facilitator in center (8 chairs if the group is small)—see the diagrams below
- **Overheads:** **A1.5,O1**—*Clapping Hands Purposes*; **A1.5,O2**—*Clapping Institute Criteria*
- **Handouts:** **A1.5,H1**—*The Meaning of Quality Summary Sheet* (1 for each person) **A1.5,H2**—*Performance Assessment—The Meaning of Quality* (1 for each person); **A1.5,H3**—*Clapping Institute Award Certificate* (1 per group)
- Optional **Overheads:** **A1.5,O3**—*Equity Discussion Questions* and **A1.5,O4**—*Special Education Discussion Questions*

Time Required:

60-90 minutes

Facilitator's Notes:

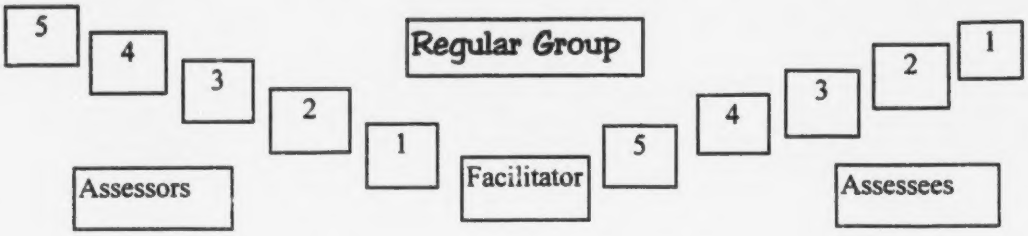
A. Set-Up (10-15 minutes)

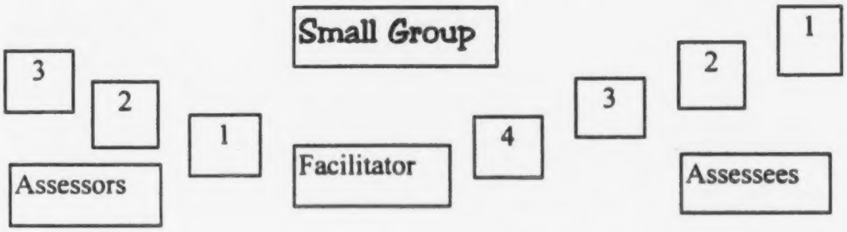
The activity is experiential—it asks participants to commit and invest themselves in the process. As such, it might make some participants uncomfortable. Therefore, it is extremely essential to manage the activity well. Keep it light-hearted.

Once this exercise has been done with a group, it can't be done again with the same group. The complete exercise requires at least 15 people. (You can get by with 11—instructions for "small group" are included where needed.)

The presenter's outline below is set up for ease of use (list style; questions to participants in bold, etc.). Typical comments made by participants at various steps of the process are included at the end.

1. Facilitator needs: paper and pencil, 1 copy of the award certificate (**Handout A1.5,H3**—*Clapping Institute Award Certificate*), and this outline.
2. Arrange 11 chairs (8 if the group is small) in a semi-circle. (See diagram below.)


Regular Group
5
4
3
2
1
Assessors
Facilitator
5
4
3
2
1
Assessees


Small Group
3
2
1
Assessors
Facilitator
4
3
2
1
Assessees

3. Introduce the exercise using **Overhead A1.5,O1**—*Clapping Hands Purposes*, as needed. Then:

   Five people (4 if the group is small) **will be asked to do a simple performance task, and five other people** (3 if the group is small) **will assess the performances. At the end everyone will be asked to describe how it felt to be in the position they took. The audience will participate at the end by questioning the volunteers on their feelings and adding their own observations.**

4. Ask for 10 volunteers (7 if the group is small) to come up and sit in the 10 (or 7) chairs arranged in a semi-circle. (Facilitator sits in the center chair—see diagram.)

5. Note names of volunteers on a seating chart

6. Point to participants to your right as the "assessees"—they will do a simple performance task

7. Point to participants to your left as the "assessors"—they will evaluate the quality of performance on the task.

BEST COPY AVAILABLE

8. Tell the volunteers:

   This activity was developed to explore the consequences of being treated in different ways; it was developed to elicit thoughts and feelings, so please be alert to them as the exercise progresses.

9. Optional ways to get more people actively involved:

   a. Ask the audience to "mirror" the panel—each person picks one volunteer and considers how they would react in the same situation.

   b. Select 1-2 math people to compute the average scores given by the assessors.

   c. After the assessors give their scores, have audience "vote" on scores with their fingers.

B. Assessment Task Administration (10-15 minutes)

1. Assessee #1 (fifth person to your right; fourth person if the group is small)

   (Name), please clap for us.
   (clap)
   Thank you, (name).

   (Give no verbal or non-verbal feedback.)

2. Assessee #2 (fourth person to your right; skip this step if the group is small)

   (Name), please clap for us.
   (clap)
   (Name), please leave the room and come back when asked.

   (Designate a "hall monitor" to ask the person to come back in when asked and to make sure the person can't hear the discussion.)

   Assessors please assess (name)'s clapping on a scale of 1-5, with 1 being low and 5 being high.

   (Record each rater's score beginning with the person on your immediate left; compute average; report to panel. Ask the participant to return, but give NO feedback.)

3. Assessee #3 (third person to your right)

   (Name), please clap for us.
   (clap)
   (Name), please leave the room and return when called.

Assessors please assess (name)'s clapping on a scale of 1-5, with 1 being low and 5 being high.

(Record each rater's score beginning with the 2nd person to your left; compute average; report to panel. Ask the assessee to return.)

OK, (name), you got a score of _______. I hope you find this information useful.

4. Assessee #4 (second person to your right)

   (Name), clap for us.
   clap)
   (Name), please leave the room and return when called.

   After the assessee leaves the room, turn to the assessors:

   I have good news and bad news. The good news is that I have heard from the International Clapping Institute. As you know, that's the organization that is developing the international standards for clapping that all our students must meet by the year 2000. They have begun a handbook on how to conduct the assessment. From now on they want us to assess all clapping performances on three traits—volume, appropriateness, and creativity. (Overhead A1.5,O2)

   The bad news is that they haven't yet finished the handbook, so they really can't give any more guidance than that. So we just have to do the best we can. So, from now on we'll assess performance on each of the three traits, where 1 is low and 5 is high.

   (Record the score for each trait—volume, appropriateness and creativity. Begin with a different assessor each time; for example, start with the 3rd person on your left for volume and the 4th person for appropriateness. Average the scores for each trait and prepare the certificate—**A1.5,H3**—*Clapping Institute Award Certificate*. Ask the assessee to come back, give him/her the certificate, and shake his/her hand.)

   (Name), here are your scores. I hope you find them useful.

5. Assessee #5 (first person to your right): This participant gets special treatment. Ask these questions:

   (Name), tell me about your previous clapping experience. In what kinds of situations do you find yourself clapping?

   You've obviously had some experience clapping. What do you feel are your strengths as a clapper?

Is there anything in your clapping you would like to improve? Anything you'd like the panel to give feedback on?

Is there any guidance that you'd like to give the assessors?

To Assessors:

You're the experts on clapping; that's why you are here. Is there anything you want (name) to know that you'll be looking for when he/she claps?

Do you want to discuss the criteria given to us by the Clapping Institute?

Is there anything you want (name) to know is important to be effective? Anything in particular you'll be looking for in the clap?

To Assessee #5:

Would you like to describe the setting in which you'd use the clap you are about to demonstrate?

What do you want to accomplish with this clap?

Are you ready? Any more discussion prior to the clap?

(Name), please clap for us.

(clap)

Do you want to leave the room or stay? Do you want feedback verbally, numerically, or both?

Assessors provide feedback as requested.

C. Debriefing (15 minutes)

1. **Each person will have the chance to describe how he or she felt during this process and the thoughts and realizations he or she had. After each individual has had a chance to talk, the floor will be open to the audience to ask questions or provide observations.**

Optional ways to get more people involved:

a. Have 2-3 recorders note key comments on chart paper during the discussion (e.g., "lacked clarity," "wanted to try again," etc.).

b. Have people "buzz" (discuss) their thoughts/feelings as mirrors before the panel reports.

2. Ask Assessee #1 (5th to your right; 4th if the group is small):

   What thoughts or feelings did you have?

   How did you feel when we came down the line and other people got other kinds of preparation or feedback?

   How did you feel about getting no preparation or feedback?

3. Ask Assessee #2 (4th to your right; skip this step if the group is small):

   What thoughts or feelings did you have?

   How did you feel when you were asked to go out of the room?

   How did you feel about getting no feedback?

4. Ask Assessors (move these questions to step 6 if the group is small):

   How did you feel about rating the second person and not the first?

   How did you feel about having no criteria?

   Did anyone give any "protest" votes? (E.g., "I'm going to give everyone a '3' or a '5' because I was not given any guidance on what to do.")

   Did anyone have the thought, "I better give myself some wiggle room by not giving the first person higher than a '3'."

5. Ask Assessee #3 (3rd to your right):

   What thoughts or feelings did you have?

   How did you feel when you were sent out of the room?

   When you got your score, was the feedback useful?

6. Ask Assessors:

   Did it bother you that I asked you to give (name-#3) his/her scores but didn't ask you to give (name-#2) his/her scores?

   Did you compare (name-#3) with (name-#2) and (name #1)?

   (Implication: when we are given no criteria to make criterion-referenced comparisons, we start comparing students to each other—norm-referenced comparisons)

7. Ask Assessee #4 (2nd to your right):

   What thoughts or feelings did you have?

   How did you feel when you came back in and got your certificate? Was the feedback useful?

   Could you hear people laughing while you waited outside? What did you think was going on?

   Did this method bother you? How did you feel about being assessed on traits you knew nothing about?

   Did you focus on any particular scores (like the lowest)?

8. Ask Assessors:

   What were your reactions when I gave you criteria categories with no further directions?

   Were these criteria useful?

9. Ask Assessee #5 (1st to your right):

   What thoughts or feelings did you have?

   Did you like the extra attention and information?

   Did you feel uncomfortable being treated differently?

10. Ask Assessors:

    Was the extra discussion useful?

    In general, did hearing other people's scores affect your scores?

11. Ask Assessees:

    Would any of you like to clap again? Why or why not?

12. Ask the audience:

    Would anyone like to make an observation about what happened or ask the volunteers questions? (Don't let this stretch on too long.)

    Has anyone been in the position of any of the assessees or assessors? Please indicate the person and relate what happened. (Keep it to 3 or 4 people.)

D. Activity "So-What." (15 minutes) (This part of the activity is essential. We want participants to be critical consumers of the performance assessments they see.)

1. Ask participants to use **Handout A1.5,H1**—*The Meaning of Quality Summary Sheet* to write what they learned from this activity about things to pay attention to when:

   - **Designing tasks**
   - **Designing performance criteria**
   - **Training raters**
   - **Preparing students**
   - **Reporting results**
   - **Considering potential sources of bias and distortion**

   [Participants will probably need "bias and distortion" defined. These are things that can mask achievement; things that can go wrong and result in drawing an incorrect conclusion about the quality of a student's work, skills, or knowledge. For example, (a) a high reading load on a math test, (b) student cultural background, or (c) student personality. See **A1.5,H2**—*Performance Assessment—The Meaning of Quality* for more examples. Note that these things don't always mess up the results. For example, if the purpose of the assessment is to assess reading comprehension, the length or difficulty of the passages is not a source of bias and distortion because that's what you're assessing. But if the amount of reading gets in the way of a student being able to demonstrate his/her math problem-solving ability, that's a source of bias and distortion.]

   Give participants 5-10 minutes to make individual notes on **Handout A1.5,H1**.

2. (5 minutes) Ask participants to note several points of agreement in table groups.

3. Make a complete list in the whole group. This list comprises criteria by which to judge the quality of performance assessments. It is the beginning of a rubric for judging quality. (Listing quality statements on chart paper enables the facilitator to post the "criteria for quality alternative assessments" to refer or add to later.)

4. Pass out **Handout A1.5,H2**—*Performance Assessment—The Meaning of Quality* to summarize what other groups have listed as important quality considerations.

Typical Responses (During Step C—Debriefing):

Assessee #1: Apprehension because I was given no guidance; every time I asked for clarification I was told, "We'll get back to you." I didn't know what was expected. I was confused. I kept wanting to try again after the criteria came out because I was sure I could do better. But, toward the end I actually felt glad I went first because by then the pressure was on. I felt it was increasingly unfair. I got angry. I felt disengaged from the process and "off the hook." It didn't matter how I did because the task was so unclear I had a good excuse.

Assessee #2: I had to guess at the criteria, so I fell back on general knowledge of what is expected in this type of situation. Therefore, personality, cultural background, and gender all might have an effect when things are unclear. I didn't know what was expected so I just tried to outdo the first person. I felt I was sent out of the room because I had done something wrong.

Assessee #3: A low score reinforced my self-judgment that I am a klutz. I was angry that I was given no rationale for my score and no opportunity to improve it. A number with no rationale or scale is not useful. A gratuitous "5" from the judge rebelling because she didn't know the criteria was not useful; it felt condescending. I would have my "folks" come in and complain to the principal. I felt uncomfortable that the first person got no score at all and the second didn't find out what his score was; it wasn't fair.

Assessee #4: I was embarrassed that I was being treated differently. I was embarrassed that some of my scores were low. I had a mental discussion with myself complaining about the exercise. I focused primarily on the low scores. I wanted to argue with the assessors. If I had known what the criteria were I could have done better. I heard people laughing and thought they were laughing at me. I was out of the room longer than anyone else so I figured I had done something wrong.

Assessee #5: The pressure was on. How could I be better than the previous ones? Knowing I would be scrutinized by my peers made me nervous. The previous people had no guidance and so could be excused for poor performance, I had no such excuse. I was embarrassed that I was treated differently. I was afraid that if I got better scores, the other "students" would be angry with me because they were treated differently. I wanted to not do a good job so that I wouldn't stand out. I appreciated the extra discussion; it made me feel more confident. Because of all the extra help, there was no excuse for not doing well.

Assessors: I was unsure of the criteria so I was afraid of giving a high score. I was angry with not having any criteria so decided to give everyone a "5". I was uncomfortable that we were asked to rate the second person and not the first. I didn't have any criteria so I was only able to rate the second performance in relationship to the first. I felt sorry for the assessees and was glad I was not one of them. I wouldn't give the first person higher than a '3' because I didn't know what the other performances would be like. Hearing the others give their scores affected my scores.

The scoring areas (traits) helped some. I'm not sure what "appropriate" or "creative" means so I made up my own idea of what each meant. I wanted to rate other things besides these three. The three categories did not capture all the relevant dimensions of performance. Not all three categories would be appropriate to assess in all contexts.

Since we knew more about the person [person #5], we held her to a higher standard than the others. Even though I specified my criteria for success, I still felt she should go beyond them to get a high score. I didn't want the assessee in the room while we gave scores because she might challenge my judgment.

Someone: One of the points invariably made is whether the clapping activity measures skills that are worth taking this much time to assess. The answer is probably "no." Clapping is just a familiar context in which to raise important points about assessment. If people have the reactions they do in this contrived situation, think about what happens in a real assessment. But, be sure to recognize that the point about "is this worth the time devoted to it" is a question we should ask of all assessments.

Variation 1: Teacher Education Faculty

A variation on the activity is to have teachers (or teacher education faculty) play the role of assessees (those being assessed) and student teachers (or students) play the role of assessors (those judging the performance). (Comment from Jamie Foster, Eisenhower Math/Science Consortium, Charleston, WV: "I did the activity immediately prior to lunch and participants discussed the activity and what they learned during lunch....The emotions that came from the teacher education faculty were insightful.")

Variation 2: Equity Enhancement

This activity brings to life with humor what is often in reality a major issue: are the assessments we use to gain portraits of student learning fair? Equitable? Do all students have the best opportunity we can provide to show their learning? Here are some additional ideas for increasing emphasis in this session on equity in assessment.

A. **Culture, Language, and Assessment.** Present the task to one of the "assessees" in a language other than English. Question the "assessee" during the debrief about how this treatment affected his/her sense of competence (not just fairness). Elicit from the observers their reactions to this unexpected switch to instructions in a language which may be completely unfamiliar, or which they might be not quite fluent in. Or prompt reflection and discussion with the question, "How does a mismatch between the language of the student and the assessment affect the quality and accuracy of the assessment?" (See the first question on **Overhead A1.5,O3**—*Equity Discussion Questions*)

Depending on the audience, this extension might include opportunities for several participants to recall similar or related assessment problems from their own experience.

Continue the discussion with the following questions on **Overhead A1.5,O3**—*Equity Discussion Questions*:

- Is this a task that is likely to draw upon experiences that ALL of your students have had? (Is clapping something that is common across cultures?)
- Are ways of clapping to communicate specific emotions the same across cultures? Do all cultures signal approval, victory, appreciation in the same ways?

- Do you think it's possible to misinterpret a student's clapping because of cultural differences?

- Obviously, we don't really want to assess clapping. It was used just to illustrate a point. However, similar things happen during real assessments. In your experience, what similar things have happened to you?

B. **Special Education Students and Assessment.** This is an assessment issue that raises strong feelings and conflicting solutions. Some people believe that students with physical, educational, or emotional challenges should be exempted from assessments—especially high-stakes assessments. Others argue that setting lower standards for some students reduces their opportunities to learn and succeed. Special educators often feel left out of professional development assessment experiences.

Select the questions below that best meet the assessment quality concerns that you wish to emphasize (**Overhead A1.5,O4**—*Special Education Discussion Questions*). (Remind the group that this activity is intended to activate and deepen their awareness of assessment quality issues that will be more fully addressed in other activities.)

- Is the clapping task one you'd use with deaf students or students with limited mobility? Why? Why not?

- Would you raise the standards for excellent clapping for students considered gifted? Why? Why not?

- Should some students be exempted from certain assessments? Why? Why not?

- Can you change the task and still assess the same learning target? What is that target anyway? Is it clapping? Could it really be about something bigger than clapping?

- Obviously, we don't really want to assess clapping. It was used just to illustrate a point. However, similar things happen during real assessments. In your experience, give some examples.

Clapping Hands

Purposes:

1. To develop guidelines for high quality performance assessment
2. To explore the importance of high quality—the impact of our assessments on students

Overhead A1.5,O1

168

Clapping Institute Criteria

Volume

Appropriateness

Creativity

169

Overhead A1.5,O2

Equity Discussion Questions

- How does a mismatch between the language of the student and the assessment affect the quality and accuracy of the assessment?
- Is this a task that is likely to draw upon experiences that ALL of your students have had? (Is clapping something that is common across cultures?)
- Are ways of clapping to communicate specific emotions the same across cultures? Do all cultures signal approval, delight, victory, appreciation in the same ways?
- Do you think it's possible to misinterpret a student's clapping because of cultural differences?
- What similar things have happened to you during real assessments that would cause ability to be mismeasured?

Overhead A1.5,O3

Special Education Discussion Questions

- Is the clapping task one you'd use with deaf students or students with limited mobility? Why? Why not?
- Would you raise the standards for excellent clapping for students considered gifted? Why? Why not?
- Should some students be exempted from certain assessments? Why? Why not?
- Can you change the task and still assess the same learning target? What is that target anyway? Is it clapping? Could it really be about something bigger than clapping?

The Meaning of Quality Summary Sheet

Task Design

Performance Criteria

Preparing Raters

Preparing Students

Reporting

Potential Sources of Bias and Distortion

Handout A1.5,H1

172

Performance Assessment—The Meaning of Quality

Things to Pay Attention to When:

Designing Tasks: be careful of public performances; treat all students the same; put the performance into a realistic context; be specific on instructions; open-ended tasks; tasks have to be meaningful to students; consider how to handle diversity and differences (e.g., special education, cultures, gender, learning styles); the task itself is a learning experience; nonthreatening; matched to valued outcomes; equity; enough time; proper resources and equipment; can be approached by the student in a variety of ways; responses can be given in a variety of ways.

Designing Performance Criteria: matched to valued outcomes; clear definitions; elaboration on how to assign points; agreement on what the criteria should be; models/examples; need rater buy-in; match performance criteria to task; range of score points; covers only the important stuff.

Preparing Raters: practice and discussion are essential; sample performances are essential; have the raters do the assessment themselves; need consistency across sites; have raters explain their ratings; calibrate the raters; practice and check consistency over time.

Preparing Students: share criteria well ahead; give students the opportunity to add to criteria; train students on what the criteria mean and how it looks when performance is good or poor; use models; do formative assessments with feedback; make the assessment purpose clear; students must have prerequisite skills; students need to trust the raters; self-assessment using the criteria is useful.

Reporting Results: make sure the scores have meaning; treat students the same; be specific and descriptive; provide evidence; allow time for discussion; emphasize what students can do, not what they can't do; nonthreatening; meaningful to students.

Avoid Bias and Distortion:

Bias in tasks, criteria or administration; criteria that don't cover the right "stuff"; poor training of raters; poor student preparation; tasks that don't elicit the right performance; sampling inadequacy; student personality; embarrassment with regard to being compared to others; changing criteria; scheduling the assessment at bad times; fatigue of raters or students; poor student or rater motivation; lack of teacher or student buy-in; rater bias; developmentally inappropriate tasks or criteria; the tendency for raters to score toward the center of the scale; examinee manipulation of the situation or raters; testwiseness; students not knowing the criteria for success; student anxiety; too much teacher help; negative teacher attitude toward the assessment; readability of assessment materials; distractions during assessment; lack of rater buy-in; cultural inappropriateness of tasks.

Handout A1.5,H2

CLAPPING INSTITUTE

AWARD

Certificate for:

____

Volume: ________________

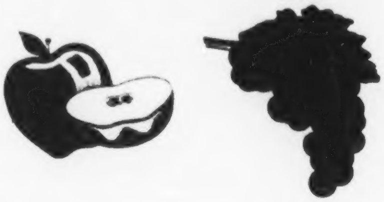
Appropriateness: __________

Creativity: __________

Handout A1.5,H3

Activity 1.6

A Comparison of Multiple-Choice and Alternative Assessment



Purposes:

1. To discuss the rationale for alternative assessment
2. To compare and contrast a traditional multiple-choice assessment with a performance assessment and discuss when each should be used
3. To think through and clarify some of the achievement targets held for students and how they match with assessment methods

Uses:

This activity is designed for those beginning their study of assessment. It can be used in **Chapter 1** to illustrate the rationale for alternative assessment and two Keys to Quality Assessment: Key 2—Clear Targets and Key 3—Target-Method Match.

Rationale:

Alternative assessments are not meant to totally replace multiple-choice tests. Rather, each has its own advantages and disadvantages, strengths and uses. This activity is intended to enable participants to construct their own interpretation of these issues. It is also a good way to begin a general discussion about assessment. Frequently it will stimulate the questions that participants have on their minds about assessment and so serves as a needs assessment. It also often results in a discussion of the rationale for alternative assessment.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Flip chart paper and colored markers

- **Overheads A1.6,O1**—*Comparison of Multiple-Choice and Alternative Assessment Purposes,* **A1.6,O2**—*Summary*

- **Handouts A1.6,H1**—*Mathematical Applications Test* (modeled after, but not taken from, any current norm-referenced, standardized tests; you could substitute a page from the 5th-grade standardized test used in your district), **A1.6,H2**—*Name the Graph*, **A1.6,H3**—*Comparison of Multiple Choice and Alternative Assessment # 1,* and **A1.6,H4**—*Comparison of Multiple Choice and Alternative Assessment #2*

- **Chapter 1** content to use as a handout (pp. 5-13)

This exercise could also be done with other grade levels and subject matter areas. Take a page from a standardized achievement test for the grade level and subject matter of interest. Then find a performance assessment with the following characteristics:

- A task that is relatively quick to read
- Generalized performance criteria that have detailed descriptions

Suggestions are provided in the **Activity 1.6 Sampler Index** (Appendix A).

Time Required:

30-60 minutes

Facilitator's Notes:

1. Use **Overhead A1.6,O1**—*Purposes* to review the purposes of the activity, as needed

2. (10-15 minutes) Give participants the multiple choice (**A1.6,H1**—*Mathematical Applications Test)* and alternative assessment (**A1.6,H2**—*Name the Graph*) examples. In small groups have them analyze their similarities and differences, and their relative advantages and disadvantages. You could have groups make two columns on a sheet of flip chart paper. In one column, have them list advantages and disadvantages (strengths and weaknesses) of the multiple-choice format; in the other, list advantages and disadvantages (strengths and weaknesses) for the performance assessment format. Have groups post their sheets.

3. (15-30 minutes) Large group discussion on the same topics. Ask participants to look at the chart pack pages posted by other groups and note similarities and differences. Summarize major points of agreement on a transparency or flip chart page (see **Overhead A1.6,O2**—*Summary).*

   (**Handout A1.6,H3** comes from a discussion with teachers, administrators, curriculum coordinators, and education assessors. **A1.6,H4** comes from a discussion of toolkit developers. They are included to illustrate the types of considerations that arise in these discussions. These can be used to assist in summarizing results or can be used as handouts.)

4. (5 minutes) The "so what" part of the activity—Make sure the following major points are made:

   - We should not totally abandon more traditional forms of assessment in favor of alternative assessment.
   - High quality assessment means having a clear notion of what we want to assess and then matching the assessment method to it.

A Comparison of Multiple-Choice and Alternative Assessment

Purposes:

1. To discuss the rationale for alternative assessment
2. To compare and contrast a traditional multiple-choice assessment with a performance assessment
3. To think through achievement targets for students and how they match with the assessment method

Overhead A1.6,O1

Summary

	Advantages	Disadvantages
Multiple Choice		
Alternative Assessment		

Overhead A1.6,O2

Mathematical Applications Test

1. John and Sue are going to the movies. If they spend $12.00 for tickets, $3.50 for popcorn and $2.50 for drinks, how much money will they have left from a $20.00 bill?

   a. $12.00

   b. $8.00

   c. $6.00

   d. $2.00

2. An advertisement states, "Of doctors surveyed, 90 percent rate no other brand of sleeping aid better than Baye's." Which of the following could you conclude from this statement?

   a. Baye's sleeping pills may or may not be better than other brands.

   b. Baye's sleeping pills are the best.

   c. Almost all the doctors in the country prefer Baye's sleeping pills.

   d. Baye's is probably worse than other brands because advertisements always lie.

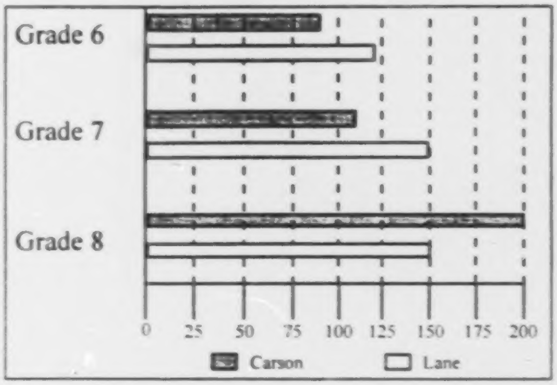
3. Quan Soo and Eddie have 9 baseball cards and 15 basketball cards. Eddie bought 6 more baseball cards. Which of the following statements is true?

   a. There are more baseball cards than basketball cards.

   b. There are more basketball cards than baseball cards.

   c. There is an equal number of basketball and baseball cards.

4. How many more eighth graders are there at Carson School than at Lane as shown on the graph below?



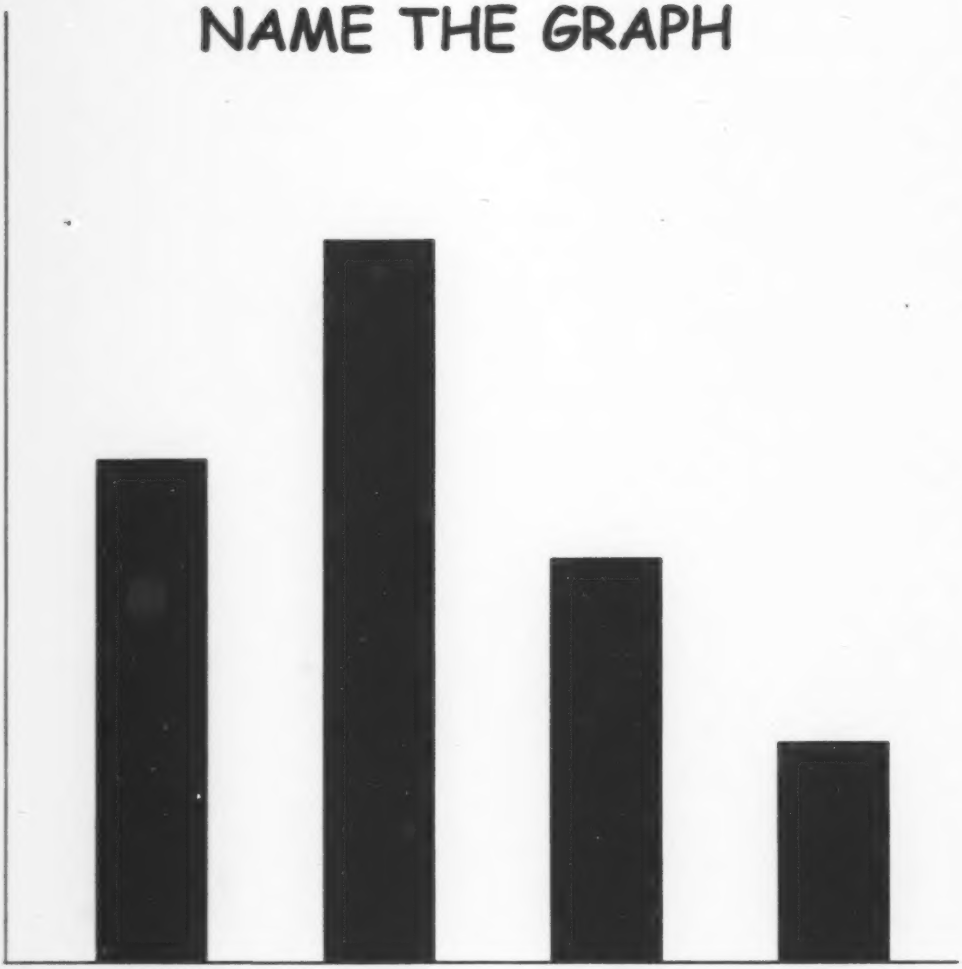
   a. 50 students

   b. 100 students

   c. 150 students

   d. 200 students

NAME THE GRAPH



1. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.

2. Write down everything you know from your graph.

Handout A1.6,H2

Comparison of Multiple Choice and Alternative Assessment #1

	Multiple Choice	Performance
Areas assessed		
• Individual process skills (e.g., interpreting graphs, addition, subtraction)	Yes	Yes
• Recall of facts	Yes	Can assess knowledge but not good for assessing the total range of knowledge for a course
• Sampling of skills and facts	Broad range	
• Understanding	Limited	Yes
• Higher-order thinking skills (e.g., critical thinking)	Limited	By design*
• Application of skills to solve real-world problems	Limited	By design*
Administrative considerations		
• Time to administer	Short to long	Usually long
• Cost-benefit	Inexpensive	More expensive
• Time and ease of scoring	Quick and easy	Slow, difficult
• Objectivity of scoring	Objective	Subjective
• Scoring criteria	Right/wrong	Rubrics
• Teacher confidence in scoring	High	Possibly low
• Parental acceptance	High	Possibly low
• Comparisons of students	Within and across classes	Usually no norms
• Accommodation of multiple learning styles	No	By design*
• Integration with school grading scheme	Maybe	By design*

*"By design" means that this is true to the extent that it is designed into the assessment.

Handout A1.6,H3, p. 1

Comparison of Multiple Choice and Alternative Assessment #1—continued

	Multiple Choice	Performance
What is communicated by the assessment		
• Importance of communication as a process skill	No	Yes
• Importance of problem-solving skills	May or may not	Yes
• Math and science as a fluid process rather than a static body of facts	No	Yes
Student concerns		
• Good reading skills	Required	Required
• Creativity	Discouraged	Encouraged
• Generation of answers	By teacher or publisher	By student
• Type of questions	Closed	Open ended
• Multiple responses	Discouraged	Encouraged
• Room for explanation	Not allowed	Encouraged
• Choice of answers	Limited; encourages guessing	No limit; guessing more difficult
• One best answer	Implied or required	By design*
• Realistic or relevant tasks	Can be	By design*
• Evidence of subject matter interconnections	Single-subject focus	By design*
• Instructs as well as assesses	No	By design*
• Appearance on paper	Intimidating	White space less intimidating

*"By design" means that this is true to the extent that it is designed into the assessment.

Handout A1.6,H3, p. 2

Comparison of Multiple Choice and Alternative Assessment #2

***Multiple-choice advantages*:** quick; good for assessing individual skills (e.g., problem solving, critical thinking) in isolation; good for assessing knowledge of facts; can sample a wide range of skills quickly; broad but not deep; cheap.

***Multiple-choice disadvantages*:** only one right answer; no opportunity for explanation; hard to measure process skills; doesn't measure ability to use skills to actually solve a real-world problem; assesses skills in isolation; might contribute to the notion that content learning is the memorization of a static body of knowledge; norm-referenced judgments of success.

***Performance assessment advantages*:** tasks can be more realistic and holistic; can see process as well as outcome; rubric can promote understanding of the target that transcends this particular problem; the task can be an instructional episode in and of itself; can fit better into a vision of assessment as a tool for learning; can assess skills and abilities hard to assess in multiple-choice format; communicates that problem solving is important; communicates that math is more than the learning of a static body of facts; communicates that math communication is important; criterion-referenced judgments of success.

Performance assessment disadvantages: can't sample as easily; more labor intensive; takes longer; more expensive; scoring is more judgmental.

Handout A1.6,H4

Activity 1.7

Target-Method Match

Purposes:

1. To examine the variety of assessment options, including various examples of alternative assessment
2. To practice matching assessment methods to the various types of learning targets we have for students
3. To realize that alternative assessment comprises many useful methods, but that there is a time and a place for more traditional methods as well

Uses:

This is an intermediate level activity that builds skills relating to **Chapter 1, Key 3**—*Target-Method Match*. Prerequisites for this activity are either (a) reading and discussing *Toolkit98* **Chapter 1**, or (b) **Activity 1.2**—*Clear Targets, What Types are These?* Before participants can match learning targets to methods, they need to be familiar with the various targets.

Rationale:

Sometimes the alternative assessment beacon becomes so bright that educators can't keep a balanced perspective on assessment. Good assessment doesn't mean using alternatives for everything; good assessment means having a clear idea of what one wants to assess (the learning targets we have for students, many times stated in local content standards) and using the method that best assesses it. This message is especially important for *Toolkit98*, which focuses on alternative assessments. We want to be clear that this focus does not imply that we recommend alternative assessments as the *sole* method to use.

But, *Toolkit98* aside, there may sometimes be pressure to use only alternative assessments. One sometimes gets the feeling that anyone using multiple-choice, for example, is evil. We advocate a more balanced approach. The key is to know *when* to use the various methods.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A1.7,O1**—*Target-Method Match Purposes;* **A1.7,O2**—*Possible Assessment Methods;* **A1.7,O3**—*Activity: Target-Method Match;* **A1.7,O4**—*Target Method Match Summary Table*
- **Handouts A1.7,H1**—*Examples of Assessment Methods;* **A1.7,H2**—*Target-Method Match;* **A1.7,H3**—*Which Assessment Method Is Best?*

Time Required:

60-90 minutes

Facilitator's Notes:

1. (5 minutes) Use Overhead **A1.7,O1**—*Target-Method Match Purposes* to introduce the activity. Good assessment means having a clear idea of what is to be assessed and then finding the best way to assess it. Sometimes multiple choice is best, sometimes performance assessment. No single method is inherently better than any other method; the goal is to match methods to targets.

2. (15-30 minutes) Use Overhead **A1.7,O2**—*Possible Assessment Methods* and Handout **A1.7,H1**—*Examples of Assessment Methods* to show the types of assessment methods. You don't usually need to discuss "selected response." Show/discuss examples of "essay," "performance assessment," and "personal oral communication," as needed. (Personal oral communication is especially important to mention because we are giving teachers permission to assess in all sorts of ways. However, the assessment, regardless of method, must be of good quality—if personal oral communication is used, it must adhere to the standards of quality described in detail in several places in *Toolkit98,* (e.g., **Chapter 3, Activity 1.5**—*Clapping Hands,* and **Activity 1.11**—*Assessment Standards.*)

3. (15-30 minutes) Break participants into groups and do the target-method match activity (as described in Handout **A1.7,H2**—*Target-Method Match* and Overhead **A1.7,O3**—*Activity: Target-Method Match*). Each group can complete a different row in the table.

   Ask each group to report their choices and why they made the choices they did. (One way to do this is by setting up a panel in

Caution

Affective targets are a red flag in some communities. If so in yours, you can do the whole activity using only the cognitive targets (knowledge, reasoning, skills and products) by removing the affective information from all overheads and handouts.

front of the group made up of a volunteer from each "learning target" group. Place X's and O's on the overhead chart [**Overhead A1.7,O4**—*Matching Achievement Targets and Assessment*] as each volunteer reports their group's work. After each "learning target," stop and discuss the similarities and differences of the responses from other groups having the same learning target.)

The completed figure should look similar to that below. (This also appears in the **Chapter 1** text.) Handout **A1.7,H2**—*Target-Method Match* summarizes the matches.

(Note: There is a fine line between "essay" and "performance assessments" that require a student to write an extended response. I like to think of "essay" as the traditional questions on tests designed to get students to elaborate on their knowledge, while "performance assessment" is based on observation and judgement. But, don't let participants get hung up on details. Sure, a writing assessment is a performance assessment based on a student essay. That's why there's an "O" on the chart in the Product/Essay cell. The important thing is that participants have a feel for how to match.)

Aligning Achievement Targets and Assessment Methods

	Selected Response	Essay	Performance Assessment	Personal/Oral Communication
Knowledge Mastery	X	X		O
Reasoning Proficiency	O	X	X	X
Skills			X	X
Products		O	X	
Dispositions/ Affect	X	O	O	X

4. (15-30 minutes) Optional challenge round: Matching quiz (Handout **A1.7,H3**—*Which Assessment Method Is Best?*). This quiz can be used for practice, to surface issues, or as a device to assess the extent of understanding of the group. (It is also fun to point out that being able to match targets to methods requires knowledge and application skills on the part of the teacher; we are assessing their knowledge with this matching quiz.)

Suggested answers are below. Remember there are no single correct answers; however, some answers are better than others. The reason that participants might disagree on the method to use for a target is that they disagree on what the target means (a major issue for assessment development, as pointed out in **Chapter 1, Key 1**—*Clear Targets* and **Activity 1.2**.) If this occurs, you might have participants generate as many meanings as they can for each target and then identify assessment methods for each.

(1) c, d, e, f. Assessing writing requires students to write; it's a skill. Writing can occur in the context of any of c-f. Portfolios (f) would be the best choice to assess the ability to write for a range of audiences and purposes.

(2) a, b, c, d, e, f, g. The method depends on which aspect of the skill is being assessed—knowledge (a, b, c) or understanding (c, d, e, f, g). Understanding could be demonstrated through an essay or through application. The important point here is that we need to be clear about what we mean by this target before we can choose the appropriate assessment method.

(3) d, e, f, g. To see whether students can contribute to groups you need to watch them do it (g); this is a skills target. Such observation can bein the context of an on-demand performance assessment or a project (d, e). The skill can be documented in a portfolio (f).

(4) a, b, c, d, e, f, g. See (2).

(5) b, d, f, g. "Demonstrates" is the key word here. The goal is not demonstrating "knowledge about" these things, but actually doing them, so it sounds more like skills and application of knowledge. Students could demonstrate understanding through a short answer (e.g., clap the rhythm) or in the context of a longer performance assessment.

(6) a, b. This seems to be at the knowledge end of the continuum; the goal is to assess prerequisite knowledge.

(7) g. These are affective outcomes. We pretty much have to watch and see this one.

(8) e. This is a product target—an extended project culminating in a written report.

(9) d, e, f, g. This is a reasoning target. Some parts of this might be assessed in selected response or short answer format, but if we really want to see whether students can bring all their skills to bear to solve problems, we need to use a performance assessment. Again, we need to be crystal clear about what we mean by the target in order to pick a best method.

(10) a, g. Here's another affective target. There are some selected response and short-answer assessments designed to get at this target. However, estimates of self-confidence can be made through watching students.

Target-Method Match

Purposes:

1. To examine the variety of assessment options, including various examples of alternative assessment
2. To practice matching assessment methods to the various types of learning targets we have for students
3. To realize that alternative assessment comprises many useful methods, but that there is a time and a place for more traditional methods as well

Overhead A1.7,O1

Possible Assessment Methods

Selected Response
- Multiple-Choice
- True/False
- Matching

Short Answer
- Fill in the blank
- Label a Diagram
- A few sentences

Essay
- Several paragraphs
- Concept map

Performance Assessments
- On-demand
- Projects
- Portfolios

Personal Oral Communication

Adapted from Rick Stiggins, Assessment Training Institute, Portland OR, 1995

Overhead A1.7,O2

Activity: Target-Method Match

Instructions:

1. Read the scenario(s) assigned to your group and answer the four questions.
2. Based on your discussion, put a large "X" in the boxes that represent the best fit of method to target type. Put an "O" in the boxes that represent a partial fit.
3. Compare your responses to those of previous groups. Agree? Disagree?

There are no single correct answers, however some answers are better than others. The important thing is to have good reasons for your choices.

Overhead A1.7,O3

Target-Method Match Summary Table

	Selected Response and Short Answer	Essay	Performance Assessment	Personal Oral Communication
Knowledge				
Reasoning				
Skills				
Products				
Affect/ Dispositions				

Overhead A1.7,O4

Examples of Assessment Methods

Sample Multiple-Choice (Selected Response) Questions

#1 (Grade 5) See *Math Applications Test* on the next page. This shows the typical kind of questions on the math applications subtest of a standardized test. (Questions 1, 3, and 4 are *knowledge*, question 2 is *reasoning*.

#2 (Grade 5) See *Going To School* that comes after the *Math Applications Test*. (This represents an attempt to assess *reasoning* using a multiple-choice format.)

Sample Short Answer Questions

#1 (Grade 3) Students have been taught how to make an anemometer and how to read a thermometer. The task is: "Each day for the next five days go outside to make your individual weather observations including temperature and wind speed and then record them on your chart." (*Knowledge*)

#2 (Grade 5) Find the average of 19, 21, 23, 25 and 27. (*Knowledge*)

#3 (Grade 6) "What is probability of getting a star when you spin the wheel?" (*Knowledge*)

#4 (Grade 9) "Use your ruler to measure the sides of the triangle that is shown. Label each of its sides with your measurements. Then draw a rectangle that has the same perimeter as the triangle. Label the sides of the rectangle that you draw." (*Knowledge*)

Sample Essay Questions

#1 (Grades 5-12) "Write a letter to a fourth grader describing how to do long division." (*Knowledge*—understanding)

#2 (Grades 6+) Describe the effect of the development of the assembly line on American society. (*Knowledge*—understanding)

#3 (Grade 9) "R, T, W, and Z represent whole numbers, and each has a different value. Use the following information to find values for R, T, W, and Z. Explain your reasoning. $R+T = R$; $(W)(Z) = Z$; $(2)(R) = (T)(R)$; $Z+W = (5)(R)$." (*Knowledge*—understanding)

Handout A1.7,H1, p. 1

BEST COPY AVAILABLE

Mathematical Applications Test

1. John and Sue are going to the movies. If they spend $12.00 for tickets, $3.50 for popcorn and $2.50 for drinks, how much money will they have left from a $20.00 bill?

   a. $12.00

   b. $8.00

   c. $6.00

   d. $2.00

2. An advertisement states, "Of doctors surveyed, 90 percent rate no other brand of sleeping aid better than Baye's." Which of the following could you conclude from this statement?

   a. Baye's sleeping pills may or may not be better than other brands.

   b. Baye's sleeping pills are the best.

   c. Almost all the doctors in the country prefer Baye's sleeping pills.

   d. Baye's is probably worse than other brands because advertisements always lie.

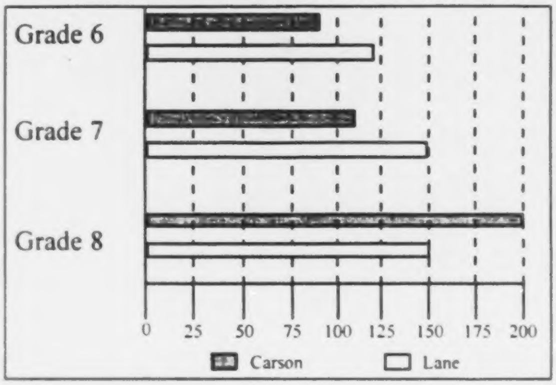
3. Quan Soo and Eddie have 9 baseball cards and 15 basketball cards. Eddie bought 6 more baseball cards. Which of the following statements is true?

   a. There are more baseball cards than basketball cards.

   b. There are more basketball cards than baseball cards.

   c. There is an equal number of basketball and baseball cards.

4. How many more eighth graders are there at Carson School than at Lane as shown on the graph below?



   a. 50 students

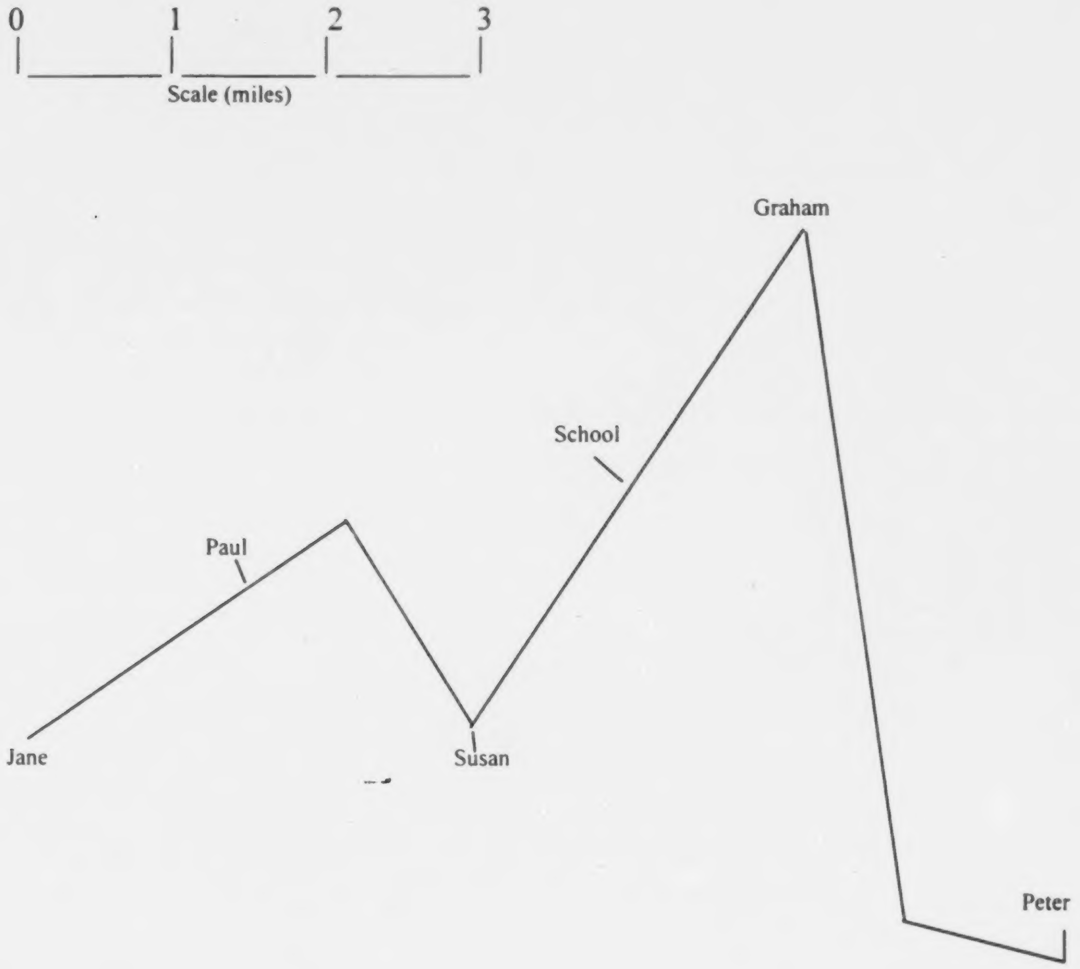
   b. 100 students

   c. 150 students

   d. 200 students

Going To School

Jane, Graham, Susan, Paul and Peter all travel to school along the same country road every morning. Peter goes in his dad's car, Jane rides her bicycle and Susan walks. The other two children vary how they travel from day to day—sometimes walking, sometimes bicycling, and sometimes going in a car. The map shows where each person lives.


0
1
2
3
Scale (miles)
Graham
School
Paul
Jane
Susan
Peter

Name: ______________________ **Date:** ____________

Going to School

1. Who lives farthest from school?
   - A. Graham
   - B. Susan
   - C. Paul
   - D. Jane

2. About how far does Susan live from school?
   - A. 1 mile
   - B. 2 miles
   - C. 3 miles
   - D. 4 miles

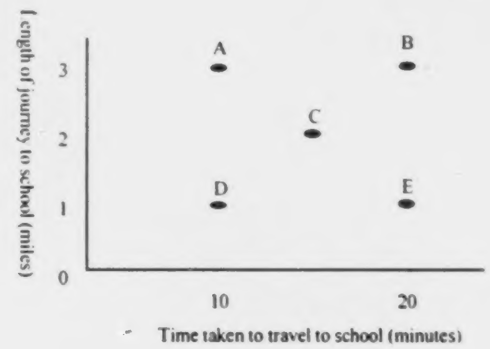
3. About how far does Jane live from school?
   - A. 1 mile
   - B. 2 miles
   - C. 3 miles
   - D. 4 miles

4. How does Peter get to school each day?
   - A. He walks.
   - B. He rides in his dad's car.
   - C. He rides his bicycle.

5. On the graph of how students got to school last Monday, who is at point A?
   - A. Graham
   - B. Jane
   - C. Paul
   - D. Peter
   - E. Susan

6. On the graph of how students got to school last Monday, who is at point C?
   - A. Graham
   - B. Jane
   - C. Paul
   - D. Peter
   - E. Susan

"How Students Got to School Last Monday"



7. If Paul and Susan both walked to school and left their houses at the same time, whom do you think would get to school first?
    A. Paul
    B. Susan
    C. They would get there at the same time.

8. On Monday, it took Jane 20 minutes to ride her bike to school. It took Susan 20 minutes to walk to school. It took Peter 10 minutes to get to school in his father's car, and Paul got to school in 10 minutes. How do you think Paul got to school?
    A. He walked.
    B. He rode his bike.
    C. He got a ride in his father's car.

9. On Monday, it took Susan 20 minutes to walk to school. It took Peter 10 minutes to ride to school with his dad. Graham got to school in 10 minutes. Which of the following is probably true?
    A. Graham walks up hill and Susan walks down hill.
    B. Graham rode his bike.
    C. Graham got a ride with his dad.
    D. Graham got lost on his way to school.

10. If it takes Susan 20 minutes to walk to school and Graham gets to school in 3 minutes, what might you conclude?
    A. Graham walked to school.
    B. Graham rode his bike to school.
    C. Graham rode to school in his father's car.
    D. It's impossible for Graham to get to school in three minutes.

From Linda Elman, Central Kitsap District, Silverdale WA, 1996.

Performance Assessment A: Sample On-Demand Tasks

#1 (Grades 3-12) Students are given a bar graph with bars only—no labels. The question to answer is: "What might this be a graph of? Label your graph and describe why you chose what you did." (*Reasoning*; *Skills*—communication)

#2 (Grade 4) Students did stage setting activities such as estimating the number of buttons in their classroom. In groups they are asked to: "Estimate the number of buttons in the whole school." (*Reasoning*—problem solving; *Skills*—communication)

#3 (Grades 5-12) In groups of five, students do the following tasks:
Part I: "Using only the streamer, find and display the average height of your group."
Part II: "Discuss your strategy so that you can describe it to others. Pick one person to be the spokesperson for your group."
Part III: "After each group shares their strategy, discuss the likelihood that their strategy would come up with the *average* height of the group."
Part IV: "What might your group have done differently to truly determine average height?" (*Knowledge*—understanding; *Reasoning*—problem solving; *Skills*—communication)

#4 (Grade 6) Students are given a table of movie ticket prices and asked: "Look at the schedule to find the price of movie tickets. Rodney is 13 years old, and both Mina and Lisa are 11. If they attend a bargain show instead of a regular show, how much money will the three children save all together? Show your work in the space below and explain your reasoning. Write your answer on the line." (*Reasoning*; *Skills*—communication)

#5 (Grades 5-12) Write about your most memorable teacher in such a way that the reader understands what makes him or her so memorable. (*Product*)

#6 (Grades 7-12) "A group of three to six students are seated in front of the class to discuss and solve a problem in a given period of time. Students are scored on strategies applied and communication skills exhibited, not on whether a solution is determined." (*Reasoning*—problem solving; *Skills*—communication and group)

#7 (Grade 9) Students circulate to six tables, each of which has a hands-on task to perform. One of the six is: "You should have a large bucket filled with beans, a small cup, magic markers, and a tray. Using any estimating strategy or population sampling techniques, estimate the number of beans in the bucket." Responses are scored on problem solving and communication in math. (*Reasoning*—problem solving; *Skills*—communication)

Performance Assessment B: Sample Assessment Projects

#1 (Grade 9) The task is: "Using only a tape measure and a stop watch, determine the amount of speeding that occurs in front of the school." Students are given stage setting instruction, and time to plan their strategy, carry out their experiments over several days, write up the results, and give an oral presentation. Students work in groups. (*Knowledge*—understanding; *Skills*—science process, group and communication; *Product*—report)

Performance Assessment C: Sample Assessment Portfolios

#1 (Grade 1) The purpose for the portfolio is to assist students to take control of their own learning. Students assemble their portfolios to tell a story about who they are mathematically. Students keep potential portfolio entries (computer printouts, real-world uses of math, journals, etc.) in a folder. About once a week students look through the folder to see what they might want to place in their portfolio. Students write why they selected each piece, how they solved a problem, reflect on what they have learned about specific topics, and use portfolios for student-led parent conferences. The portfolios are assessed on the degree to which students have presented a coherent view of themselves as learners. (*Product*; *Skills*—self-reflection)

#2 (Grades 4 and 8) Students select 10-20 items for their portfolios to demonstrate their ability to solve mathematical problems. They include 5-7 best pieces including 1 puzzle, 1 investigation, and 1 application. Other than this, they are free to choose any work in or out of school. They cannot select any more than 2 pieces of group work. They write a letter to the evaluator explaining why they choose the pieces they did and what they show about themselves as problem solvers. Portfolios are scored for problem solving and communication in math. (*Reasoning*—problem solving; *Skills*—communication)

#3 (Grades 10-12) The portfolio is used as one piece of information in deciding on an advanced endorsement in science on high school diplomas in California. During a year of science (biology, chemistry, physics, etc.) students assemble their portfolios which must contain three pieces: a problem solving investigation; a creative expression of understanding of a science concept (model, game, video, art, poetry, etc.); and growth through writing (written demonstrations over time showing how learning on a topic has evolved). Each piece is accompanied by a written self-reflection explaining why it was selected. Rubrics for evaluating the quality of each piece are included. (*Knowledge*—understanding; *Reasoning*—connections; *Products*—portfolio, research report, creative expression; *Skills*—science process, communication group)

Handout A1.7,H1, p. 7

Personal Oral Communication

#1 (Any grade level) Teachers ask students questions and listen to the quality of their responses. (*Knowledge*—understanding; *Reasoning*)

#2 (Any grade level) Teachers conference with students about their written work. (*Knowledge*—understanding; *Reasoning*; *Product*)

#3 (Any grade level) Teachers listen to students during group work. (*Knowledge*—understanding; *Reasoning*; *Skill*—group)

#4 (Any grade level) Oral student presentations. (*Skill*—communication)

Handout A1.7,H1, p. 8

Target-Method Match Activity

Instructions:

1. Attached are five scenarios, each relating to a different type of student learning target: content knowledge, skills, reasoning, products, and affect (dispositions). Read through the scenario(s) assigned to your group and answer the four questions.
2. Based on your discussion, put a large "X" in the boxes that represent the best fit of method to target type. Put an "O" in the boxes that represent a partial fit.
3. Compare your responses to those elaborated by previous groups. Where do you agree? Disagree?

There are no single correct answers, however some answers are better than others. The important thing is to have good reasons for your choices.

	Selected Response & Short Answer	Essay	Performance Assessment	Personal Oral Communication
Knowledge				
Reasoning				
Skills				
Products				
Affect/ Dispositions				

Handout A1.7,H2, p. 1

Assessing Student Mastery of *Content Knowledge*

Scenario: Assume you want your students to master some specific subject matter knowledge because it represents an important foundation for later work. To reach this goal, you plan a series of instructional activities to help your students master this required material. Now you want to assess to be sure they've got it. In this particular case, you want them to know the material outright, not through the use of reference materials.

Question #1: Should you assess mastery of this material using selected response or short answer modes of assessment, such as multiple choice, true-false, or matching exercises? Yes or No. Please briefly explain your response.

Question #2: Should you assess your students' mastery of this material using an essay form of assessment. Why or why not? Please write a brief defense.

Question #3: If you wanted to assess students' mastery of content knowledge as specified above, should you use a performance assessment? Yes or No. Defend your answer.

Question #4: Do you think the personal oral communication form of assessment—for example, by oral exam, interview, conference, or discussion—could provide you with a viable assessment of your students' mastery of this content knowledge? Why or why not?

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 2

Check Your Responses to the *Content Knowledge* Mastery Scenario

Question #1: If you answered, "Yes," to the query about assessing mastery of content targets using selected-response or short answer assessment formats, you were correct. These kinds of test items can sample student mastery of knowledge very effectively and efficiently. But remember, they tend to focus on mastery of individual concepts and facts and typically pay less attention to student mastery of the relationships among elements of knowledge.

***Question #2*:** The match between knowledge targets and the use of the essay form of assessment can be a strong one, too. Further, this match permits a careful analysis of the learner's mastery of the relationships among the parts. But remember, since it takes far more time to respond to each exercise, we can include fewer essay questions than selected response items in the same amount of testing time. So the scope of the coverage is narrowed.

***Question #3*:** The performance assessment option is far trickier in assessing knowledge mastery as a distinct target. In one sense, it appears that performance assessment could provide a quality match. If we give students a performance exercise that requires the creation of a complex product and the performer does an outstanding job, we have in hand compelling evidence that the student has (a) mastered the requisite knowledge, (b) reasoned productively, and (c) performed required steps in a skillful manner—all building blocks of the product development competence demonstrated. "Well, then," you may be saying, "that means there is a strong match between the knowledge target and the performance assessment method." Not necessarily.

The problem with using performance assessment to assess knowledge mastery becomes apparent in those instances in which the student fails to perform adequately on the complex performance assessment. When this happens, all we can see is the inadequate product. But we don't know why the performer failed. Was it due to a lack of knowledge, a failure to reason, or a failure to carry out steps in a skillful manner? We cannot know the answer to this question unless and until we use one of the other three assessment methods in a follow-up assessment to determine if the prerequisite knowledge is present or absent. So, if our goal was to assess mastery of that knowledge, why not simply start with one of those other methods? Why waste time on the performance assessment? Also, each performance assessment takes longer. Covering *all* important content using a series of performance assessments is not efficient.

With this kind of target, performance assessment is just too complex. We can choose from three better options, each of which is more focused and more efficient.

Question #4: Yes, we can assess mastery of content knowledge using personal oral communication—we can ask content questions and listen to the respondent's answers. While this might not be the most efficient way to assess mastery of content, it is an option.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 3

Assessing Mastery of *Skills*

Scenario: Although assessing reasoning proficiency is next on our list, let's hold on to that for just a bit and focus instead on skills outcomes. This will permit us to return to reasoning targets from a more informed perspective in Scenario #3.

In this scenario, assume that you teach French and wish to assess your students' skill at communicating in that language in a conversational situation. So the skill of *oral language proficiency* is your target.

Question #1: Can you assess oral language proficiency in a conversational context using a selected response or short answer mode of assessment? Why or why not? Defend your answer.

Question #2: Can you assess these skills using an essay form of assessment? Yes or No. Why?

Question #3: Will performance assessment work as a basis for assessing the foreign language speaking proficiency of your students? Why?

Question #4: Can you use personal oral communication as a basis for assessing conversational skill in a second language? Defend your response.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 4

Check Your Responses to the *Skill* Mastery Scenario

Question #1: If the objective is to assess ability to demonstrate the actual skill, we cannot do it with a selected response or short answer method of assessment. Remember, we are dealing with speaking and listening competence here. Reading and writing are not going to work. But don't overlook the fact that we can use the selected response or short answer methods to assess: (1) student mastery of some of the prerequisites of effective communication, such as vocabulary; or even (2) knowledge about important things to remember when, say, giving an oral presentation.

Question #2: In this case, the same is true of the essay method of assessment. We can't assess speaking and listening competence with a written assessment, but we can get at some of the important prerequisites.

Question #3: Here is where we begin to find the good matches between assessment method and target. Yes, we can use performance assessment to reflect conversational skill in a second language. For example, we can have two students converse, observe, and listen to the interaction. To make this work, we must have clear criteria in mind—what does "conversational skill" look like? What are the features you'll look for? This is required in order to make dependable judgments about proficiency.

Question #4: We get a similarly strong match when we use personal oral communication assessment methods in cases such as this, when oral communication proficiency is the target. Of course, you realize that this works best when you, the teacher, know the language too! And, you'll need high quality performance criteria.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Assessing *Reasoning* Proficiency

Scenario: Now let's turn to reasoning targets, evaluating matches with various kinds of assessment methods. In this case, you are a teacher who has seen to it that your students are able to access important knowledge when required. Now you want to see if they can use that knowledge productively to solve relevant problems. You want to see if they can reason analytically (think about the parts of things) and comparatively (think in terms of similarities and differences), draw inferences, and think critically (take and defend a position on an issue, for example).

Question #1: Can you get at these things with selected response or short answer assessments? Explain why or why not.

Question #2: Does the essay method of assessment work in contexts where we seek to assess reasoning proficiency? Why or why not?

Question #3: Is performance assessment a viable alternative when assessing reasoning proficiency? What do you think? Why or why not?

Question #4: Can we use personal oral communication as an assessment method to probe a student's ability to use knowledge to reason effectively and solve problems? Defend your response.

Application Question: Let's say you want to assess your students' ability to use a certain knowledge base to reason productively, but you are not sure all of the students have had a chance to master that important knowledge. Could you still assess their reasoning proficiency in this context or would you have to abandon your assessment? Under what conditions, if any, could you proceed?

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 6

Check Your Responses to the Assessing *Reasoning* Scenario

Question #1: Can you use selected response or short answer for assessing reasoning targets? You bet—but not for all kinds of reasoning. We can pose multiple choice questions, for example, that ask students to select the best analysis from among several, identify an important similarity or difference, or make the correct inference from among several possibilities.

But to make this work, three critical conditions must be satisfied. The assessor must (1) start with a clear sense of the kinds of reasoning to be assessed (analytical, comparative, etc.), (2) know which of these valued targets can be translated into selected response methods of assessment (some cannot), and (3) know how to write test questions that tap different kinds of reasoning and result in only *one* correct answer. Also remember that using selected response assesses reasoning skills in *isolation*. If you want to know whether students know *when* to use a skill, or *how* to use skills together to actually *do* something, another format is required. So, selected response and short answer are partial matches.

Question #2: The essay mode of assessment can work very well for assessing reasoning when students are asked to describe their reasoning processes in their responses. But to make this method work, there is a condition to be satisfied. We must evaluate student essays using scoring criteria that include a focus on effective reasoning. For instance, if students are to write a comparative essay in response to an exercise, what does a good comparison look like? What are the active ingredients? These need to be reflected in the scoring criteria.

Question #3: Yes, performance assessment can work too. We can present students with an exercise that asks them to demonstrate the skillful performance of reasoning—for instance, designing a science experiment. Or we can ask them to create a complex achievement-related product such as a comprehensive laboratory research report, and we can apply evaluative criteria reflective of proper reasoning as we evaluate that work. But again, a clear set of reasoning targets, sound exercises, and appropriate criteria are all essential.

Question #4: This makes four for four! Personal oral communication can work well for assessing reasoning. And one of the great strengths of this option is the fact that we can ask follow-up questions to examine more carefully the kinds of reasoning that underpin student responses to our problem-solving challenges. Of course, appropriate criteria for judging reasoning are essential.

Application Question: We can proceed in the absence of assurance that all examinees have had access to the prerequisite knowledge if we give them that knowledge and then ask them to use it. One way to do this is to offer an open-book test. Another is to provide a map, chart, or brief passage of material along with the assessment exercises and then ask respondents to use this material to answer the questions. A third way is to have students work in groups to share or generate important prerequisite knowledge before individually attempting the reasoning task.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 7

Assessing the Ability to Create Quality *Products*

Scenario: You want your students to be able to create quality products—products that meet certain specified standards. They might be samples of writing, term papers, technology products, craft products, artistic creations, or others. Your instruction has centered on helping students learn the differences between products that are of high and low quality. You have provided practice in developing products that meet your standards. Now it is time to assess the students' achievement to see if your instruction was effective.

Question #1: Can you assess the ability to create these kinds of products using selected response or short answer modes of assessment? Why or why not?

Question #2: Will essay assessment work for evaluating this kind of achievement? Explain.

Question #3: Can performance assessment provide the evidence of proficiency needed to evaluate this kind of achievement target? Defend your response.

Question #4: Is personal oral communication a viable way to assess when products serve as the source of evidence of proficiency? Yes or no. Why?

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p. 8

Check Your Responses to the *Product* Development Scenario

Questions 1-4: Of the four, only one method can work in this case: performance assessment. Students must create the actual product for observation and evaluation. But that does not mean that there is no role for the other three in contexts in which learners are striving to become proficient at creating such products. Selected response, essay, and personal communication methods can help us to determine if students are mastering the knowledge that is a prerequisite to such creation. This might include knowledge of the attributes of quality products or of the steps to be taken to produce quality products.

Here again, there must be clear and appropriate performance criteria for judging the quality of the resulting products.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Assessing *Affect* and *Dispositions*

Scenario: Now let's turn to affective targets. In this case, you are a teacher who has seen to it that your students are able to access important knowledge when required; you've determined that they possess the knowledge and skills needed to do quality work. Now you want to see if they have certain habits of mind and other organization skills that will enable them to *want* to use what they know when it's needed. Such affective and dispositional skills include: attitude toward school or a specific subject area; self-confidence; perseverance; and motivation.

Question #1: Can you get at these things with selected response or short answer assessments? Explain why or why not.

Question #2: Does the essay method of assessment work in contexts where we seek to assess affect? Why or why not?

Question #3: Is performance assessment a viable alternative when assessing affect? What do you think? Why or why not?

Question #4: Can we use personal oral communication as an assessment method to probe a student's dispositions? Defend your response.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p.10

Check Your Responses to the Assessing *Affect/Dispositions* Scenario

Question #1: Can you use selected response or short answer for assessing affective targets? You bet—there are many surveys and self-rating forms for looking at dispositions. Remember, though, that these are all self-report—the student is answering for him or herself; and sometimes students answer in ways that are socially acceptable or that they think the assessor wants to hear.

Question #2: The essay mode of assessment can work for assessing dispositions. In this case students can write about, for example, their attitude toward reading. Remember, however, that the assessor will have to interpret what the student says to draw conclusions about affect; so criteria for desirable responses must be developed. In other words, what do responses from highly motivated students look like?

Question #3: Yes, performance assessment can work too. We can engage a student in a task and look for evidence of perseverance, attitude, etc. But again, appropriate criteria for judging performance are essential. What does it look like when a student is self-confidently pursuing a task?

Question #4: This makes four for four! Personal oral communication can work well for assessing affect. And one of the great strengths of this option is the fact that we can ask follow-up questions to clarify what students are telling us. Here, again, we need criteria for judging and interpreting what students say.

Caution: As everyone is well aware, the affective domain represents student learning targets that are very sensitive. In fact, some states and districts leave affective targets out of official content standards. We must be very clear that what we assess is achievement related and that everyone agrees on the importance and definition of the target. We must also be very careful that how we assess such targets is non-invasive.

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Handout A1.7,H2, p.11

Matching Achievement Targets and Assessment Methods

Achievement Target	Selected Response Short Answer	Essay	Performance Assessment	Personal Oral Communication
MASTERY OF KNOWLEDGE	Can be used to assess content knowledge	Can serve to assess student mastery of complex structures of knowledge	Cannot be used as efficiently as the other three methods	Can assess small domains of knowledge when short-term record keeping is required
REASONING	Can be an excellent way to assess *some* kinds of reasoning in isolation	Can provide a window into reasoning	Strong match—products or performances can be observed for reasoning	Can ask student to "think out loud" to examine reasoning proficiency
SKILLS	Can only test for mastery of simple prerequisite procedural knowledge	Can ask students to describe complex "how to" procedural knowledge. Good match if the skill being assessed is writing.	Strong match—can observe and evaluate skills as demonstrated	A strong match when the skill is oral communication proficiency; can ask student to describe and discuss complex "how to" procedural knowledge
PRODUCTS	Can only test for prerequisite knowledge of the attributes of quality products	Can only tap ability to describe the attributes of a quality product, unless the product being assessed includes writing.	Strong match—can assess (a) proficiency in carrying out the steps required to create a quality product, and (b) the product itself	Can probe procedural knowledge and/or knowledge of attributes of a quality product

Adapted from Rick Stiggins, Assessment Training Institute, Portland Oregon, 1995

Which Assessment Method Is Best?

Instructions: Match the learning targets on the right to the assessment method on the left. There are no single correct answers, however, some responses are better than others. The important thing is to explain *why* you made the match you did.

Assessment Method	Student Learning Targets
a. Selected response (multiple-choice, matching, true-false)	1. Writes effectively for various audiences and purposes
b. Short answer (fill in blank, a few sentences, label diagram,)	2. Develops knowledge and understanding of how language is used
c. Essay	3. Contributes productively to both large and small groups
d. On-demand performance assessment	4. Uses historical and contemporary information to demonstrate cross-cultural understanding
e. Performance assessment—project	5. Demonstrates knowledge of melody, rhythm, harmony, pitch, form, timbre, and tone quality
f. Performance assessment—portfolio	6. Uses music vocabulary correctly
g. Personal oral communication	7. Exhibits self-confidence, self-discipline, and self-motivation
	8. Develops a personal wellness plan that recognizes the impact of nutrition and fitness on personal well-being
	9. Chooses and devises appropriate strategies to solve mathematical problems
	10. Views self as capable of using mathematical skills and knowledge to make sense of new problem situations.

Handout A1.7,H3

Activity 1.8

Sam's Story: Comprehensive Assessment*

here's
my
best
work!

Purposes:

1. To understand the importance of using multiple measures to get a more complete picture of a student
2. To understand that each form of assessment has strengths and weaknesses
3. To understand the need for alignment of outcomes, instruction, and assessments (we must match student achievement targets to assessment methods)
4. To understand the concepts of "sampling,""target-method match," matching purposes to assessment plans, and "validity"

Uses:

This is a good introductory activity to use in **Chapters 1** or **3** to illustrate the ideas of "sampling," "validity" (bias and distortion), "target-method match," and alignment between assessment and instruction.

Rationale:

Participants may have heard rhetoric that suggests that one form of assessment is better than another in all circumstances and for all purposes, and that a single assessment should yield all the information needed. Actually, no single type of assessment can yield all the information needed for all users and uses. And, different types of assessment methods are better suited to gather information about different types of student achievement targets (see **Chapter 1, Key 3**—*Target-Method Match*) or for different purposes (see **Key 2**—*Clear and Appropriate Purposes)*. In this activity, participants will gain perspective on the strengths and weaknesses of various types of assessments.

---

* Based on an activity by Carol Perlman, Chicago Public Schools

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens; or a flip chart and pen
- **Overheads A1.8,O1**—*Sam's Story Purposes* and **A1.8,O2**—*Sam's Story Summary Chart*
- **Handout A1.8,H1**—*Sam's Story* for each participant

Time Required:

40-60 minutes

Facilitators Notes:

1. As needed, use **A1.8,O1**—*Sam's Story Purposes* to introduce the activity.
2. (5-10 minutes) Ask participants to read the handout (**A1.8,H1**—*Sam's Story*) and complete the chart individually. This task requires participants to rate how much they would trust each piece of assessment evidence as:

    - A sole measure
    - One of several highly relevant measures
    - Not directly relevant, but might be useful
    - Wouldn't tell much that's useful

    Ask participants to note the reasons for their choices.

3. (5 minutes) When participants are done, reassemble the group. For each of the 10 items, ask participants to hold up their hands to indicate how they'd trust each item. Record numbers in each box. This helps to illustrate the variability of opinion based on knowledge and experience.

    Indicate participant responses on the Overhead **A1.8,O2**—*Sam's Story Summary Chart* by drawing a line though the category(ies) that represent the majority of responses.

4. (10 minutes) After going through all 10 items, point out (if the case) that there's nothing they'd trust as a sole measure of Sam's performance. Why is that? (Answers may include: need to know what each assessment is measuring, need to know how the assessment relates to the class curriculum, need more examples of Sam's work in math, the different tests may be measuring different things.)

Look at the measures that were rated the best. What do these have in common? (Possible answers: they're all closely related to the target that we were making a decision about; they include multiple samples of work; they are "authentic.")

Look at the measures that were rated the lowest. Why did they get such low ratings? (Possible answers: they may be biased, marginally relevant to the decision at hand, not "authentic" enough, or not comprehensive enough.)

5. (20 minutes) Note that appropriate educational decisions depend on the accuracy and comprehensiveness of educational assessment. Inaccurate or partial information may cause harm. Ask participants for examples of how inaccurate or partial information may cause harm to students.

Tie to the concept of *sampling* (**Chapter 1**, **Key 4**—*Bias and Distortion)*: *When we try to determine Sam's performance in math, it's obviously impractical to see how well he can solve every possible problem there is. Nonetheless, you as a teacher want to get an accurate idea of how well Sam can handle the full array of possible tasks that Sam might be expected to encounter. That full array of possible tasks is called a "domain of knowledge."*

Since it's impossible to see how well Sam can perform on the entire range of math problems, we have to select a "sample" of math problems from the entire assessment domain. We try to sample the contents of a domain in a representative fashion so that, based on how well students do on the sample, we can make inferences about the students' status with respect to the entire domain.

Ask participants which items on the list "sample" best. (Answers: in-class math tests, portfolio.)

Tie to the concept of *validity* (**Chapter 1**, **Key 4**—*Eliminating Sources of Bias and Distortion*): Validity has to do with the extent to which we can make good decisions such as:

- Are we looking at the right information?
- Are we measuring the right thing?
- Are we paying attention to things that aren't really relevant?
- Are we inadvertently measuring something else?

Ask participants which items on the list they feel are the most valid for the purpose of making a mastery decision? Why?

Tie to the concept of *matching method to purpose* (**Chapter 2**, **Key 2**—*Clear and Appropriate Purposes)*: *Note that the reason the teacher wanted information about Sam was to "decide about Sam's mastery of the content and to design an instructional plan." Ask teachers if one's decisions about which evidence to trust would be different if the* *purpose were different, for example, to decide if Sam were working up to potential? In that case, what measures might be best?*

6. (10 minutes) Conclude the discussion by asking the following questions of the group. To make a decision about Sam's mastery of math and to design an instructional plan for him, what additional information do you need to know:

   - About Sam?
   - About his last math teacher and that teacher's course content, instructional strategies, and grading criteria?
   - About the tests used to provide these results?

Sam's Story: Comprehensive Assessment

Purposes:

1. To understand the importance of using multiple measures to get a more complete picture of a student
2. To understand that each form of assessment has strengths and weaknesses
3. To understand the need for alignment of outcomes, instruction, and assessments
4. To understand the concepts of: sampling, validity and target-method match

Sam's Story Summary Chart

here's my best work!	Would trust as a sole measure	Would trust as one of several relevant measures	Not directly relevant, but information might be useful	Wouldn't tell me much that's useful
1. In-class math tests from last year				
2. IQ score				
3. Norm-referenced, multiple choice math test				
4. State criterion-referenced test of mathematical skills				
5. One example of problem-solving ability on a math word problem				
6. Psychological assessment results				
7. Portfolio of work in math				
8. Self-report on performance in math				
9. Sam's teacher's assessment of his performance in math				
10. Sam's parents' assessment of his performance in math				

Overhead A1.8,O2

Sam's Story

You are Sam's new teacher for the coming school year and are trying to determine Sam's level of proficiency in math. You are interested in making a decision about Sam's mastery of the mathematics content taught and to design an instructional plan for him. How much faith would you place in each of the following sources of information?

	Would trust as a sole measure	Would trust as one of several highly relevant measures	Not directly relevant, but information might be useful	Wouldn't tell much that's useful
1. Sam's math scores on his in-class math tests from last year				
2. Sam's IQ score				
3. Sam's math scores on a norm-referenced, multiple choice test				
4. Sam's math scores on a state criterion-referenced test of mathematical skills				
5. One example of Sam's problem-solving ability on a math word problem				
6. Sam's psychological assessment results				
7. A portfolio of Sam's work in math				
8. Sam's self-report on his performance in math				
9. Sam's teacher's assessment of his performance in math				
10. Sam's parents' assessment of his performance in math				

Handout A1.8,H1

Activity 1.9

Going to School

Purposes:

1. To understand the importance of clearly defined and well thought-out performance criteria
2. To understand different types of performance criteria
3. To discuss the relative advantages and disadvantages of different types of performance criteria

Uses:

Part A of this activity can be used in **Chapters 1** or **2** to illustrate the importance of performance criteria as one of two parts of an alternative assessment. Part A is an intermediate level activity. Prerequisites might be: **Activities 1.1**—*Changing Assessment Practices, What Difference Does it Make For Students?* or **1.12**—*Assessment Principles*; and **Activity 1.5**—*Clapping Hands*.

In **Chapter 3**, Parts B and C can be used to illustrate types of performance criteria and discuss the advantages and disadvantages of each kind. Parts B and C address advanced topics. Prerequisites, in addition to those listed above, might be: **Chapter 2** text and **Activity 2.1**—*Sorting Student Work*.

Rationale:

Participants frequently need help in defining "performance criteria," understanding why they are important, seeing the difference between tasks and criteria, understanding the role of performance criteria in instruction and assessment, and understanding the differences between types of criteria and their advantages and disadvantages. This activity is designed as a hands-on way to explore these topics.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads: A1.9,O1**—*Going to School Purposes,* **A1.9,O2**—*Overhead of Graph,* **A1.9,O3**—*Holistic/Analytical Trait,* **A1.9,O4**—*Generalized/Task Specific,* **A1.9,O5**—*Quote on General Criteria,* **A1.9,O6**—*Rubric's Cube*
- **Handouts: A1.9,H1**—*Going to School,* **A1.9,H2**—*Devon and Mickie,* **A1.9,H3**—*Task Specific Scoring Rubric,* **A1.9,H4**—*Task Specific Scoring Guide #2,* and **A1.9,H5**—*Generalized Scoring Rubric—Relative Quality.* Add to this two other rubrics from the **Sampler** in **Appendix A**. Suggestions are provided in the **Activity 1.9 Index** in **Appendix A**.
- Extra assessment samples if Part C is done—see the **Activity 1.9 Index** in the **Appendix A Sampler** for suggestions.

Time Required:

1¼ to 2 hours

Facilitator's Notes:

Use **Overhead A1.9,O1** to describe the purposes of the activity.

A. Importance of Performance Criteria (30-45 minutes)

1. (10 minutes) Hand out the *Going to School* (**A1.9,H1**) problem and have participants solve it in small groups

   Use **A1.9,O2**—*Overhead of Graph* to elicit and show the right answer.

   (Note: If participants come up with odd answers or say they couldn't figure it out because the problem was incomplete or confusing, use this as an opportunity to critique the task. Reinforce the notion of "sources of bias and distortion" (**Chapter 1, Key 4**) by pointing out that confusing tasks may mask student ability to demonstrate that they can problem solve. Ask participants how they would make the task better.)

2. (5-10 minutes) Hand out the two student solutions (**A1.9,H2**) and ask participants to "grade them" or "score them" individually. Participants might ask questions like, "Do you want us to use A-F?" "Do we use a four-point scale?" "Do we take spelling into account?" etc. Just tell them to use their usual procedure. When individuals are done, ask small groups to discuss the various procedures used and any problems they ran into.

3. (5-10 minutes). Reconvene the whole group and ask 4-5 individuals from various small groups to discuss the difficulties they encountered. Their comments will illustrate: (a) the random noise or error that can affect scoring when clear criteria are not in place, and (b) without a definition of how and what to score, it is hard to interpret what the grades or scores mean.

4. (5 minutes) Pass out the criteria packet (**A1.9,H3; A1.9,H4; A1.9,H5;** plus two samples from **Appendix A**). Ask the participants to score the student response with **A1.9,H3** and **one sample from Appendix A**.

5. (5-10 minutes) Ask small groups to discuss the extent to which scoring was easier and more consistent when they had performance criteria to use. Make sure the following points are made. Rubrics are needed to do the following:

   a. Make it clear to students what it takes to be successful

   b. Make it clear to raters how to assign scores in a consistent manner

B. Optional Activity—Different Types of Performance Criteria (45-60 minutes)

1. (5-10 minutes) Ask participants to score the student responses with the various rubrics in this activity and those chosen from Appendix A. Begin with "holistic, skimpy" and progress to "analytical trait, detailed." Otherwise, experience with more detailed schemes will influence the skimpier versions. This procedure also illustrates how more detail in a rubric can lead to more accurate scoring.

2. (15 minutes) For each rubric, have participants identify the following.

   a. Task-specific or general (they should get this right—it's stated on the top of each rubric): **A1.9,H3** and **H4** are "task-specific." The others are general. Ask participants to describe to each other the difference between task-specific and general criteria.

   b. Analytical trait (more than 1 score) or holistic (1 score). **A1.9,H3** and **A1.9,H5** are holistic; **A1.9,H4** is analytical trait. One of your examples from **Appendix A** is holistic; the other is analytical trait. Ask participants to describe to each other the differences.

   c. Detailed or skimpy. **A1.9,H5** is skimpy—terms are not defined well. Discuss the impact of skimpiness on consistency in scoring between raters and the ability of criteria to communicate to students what it takes to be successful. Both samples from **Appendix A** are detailed. Ask participants if detail would help students self-assess.

3. (20-30 minutes) Ask participants to find examples in **Appendix A**—*Sampler* of the following types of performance criteria (a single sample may illustrate several features):

- Holistic
- Analytical trait
- General
- Task specific
- Skimpy
- Detailed

Participants can use the "Rubrics Cube" (**Overhead A1.9,O6**) to classify the samples they find.

C. **Optional Activity**—Advantages and Disadvantages of Different Types of Performance Criteria (20-30 minutes)

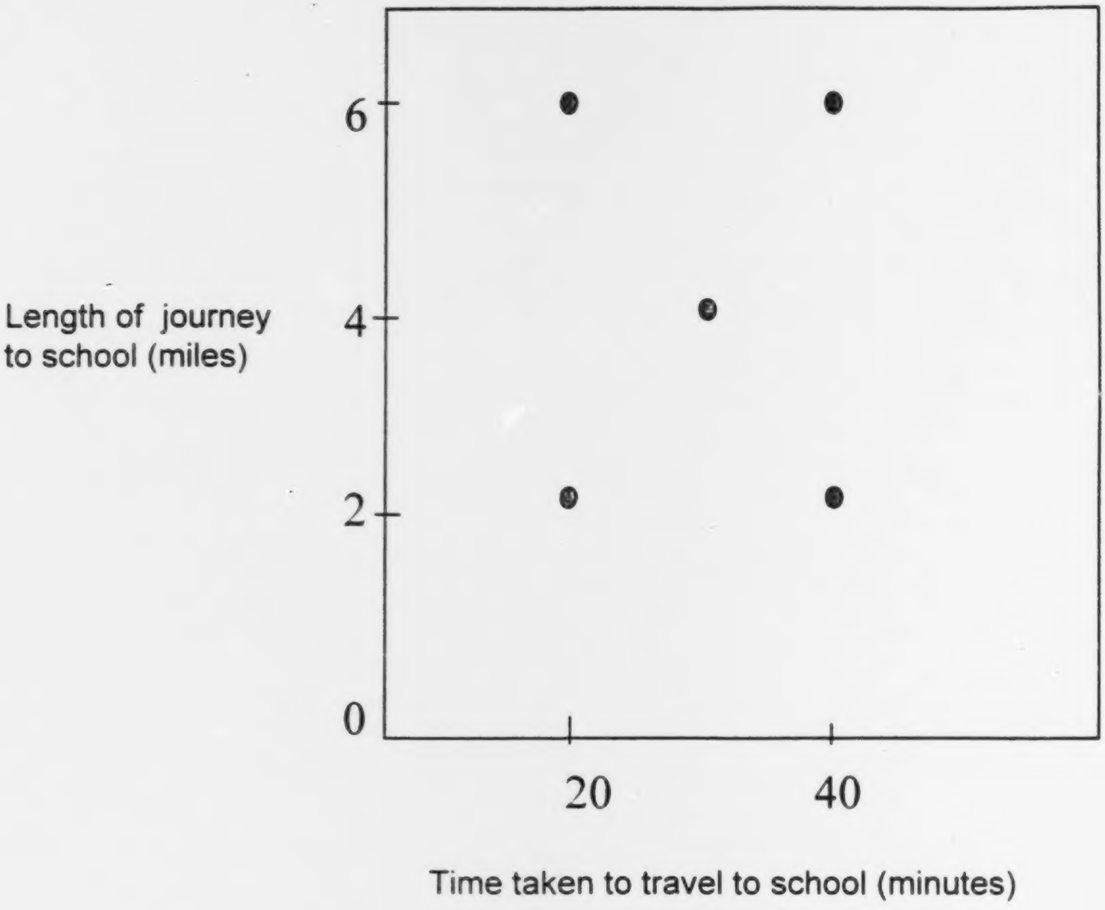
Note: Don't do this part of the exercise if you don't feel you can explain when to use each type of rubric/criteria.

1. (20-30 minutes) Discuss the relative advantages and disadvantages of the different types—small group and then large group. Advantages and disadvantages of each method are summarized on **A1.9,O3** and **A1.9,O4**. Ask participants to add to or clarify what's on the overheads. Participants can also use the "Rubrics Cube" (**Overhead A1.9,O6**) to list the strengths and weaknesses of each type of rubric.
2. (5-10 minutes) Bottom line—analytical trait, general criteria are better for instruction. The quote on **A1.9,O5** is from a major national study of the impact of performance assessment on teaching and learning. Ask participants to explain to each other why general criteria are better instructionally.

Going to School

Purposes:

1. To understand the importance of clearly defined and well thought-out performance criteria
2. To understand different types of performance criteria
3. To stimulate discussion about the relative advantages and disadvantages of different types of performance criteria



Overhead A1.9,O2

Holistic	Analytical Trait
Definition: A single overall score is assigned to a student performance or product	Definition: More than one dimensions or traits of a student performance or product are scored
+ Easy to train raters	- Harder to train raters
+ Quick scoring process	- Slower scoring process
+/- Assesses overall student performance	+ Shows the relative strengths and weaknesses of student work: this helps both teachers and students
- Two students can get the same score for different reasons	+ Reasons for differences in scores are more readily apparent

Overhead A1.9,O3

Generalized	Task Specific
Definition: Criteria can be used across tasks	Definition: Criteria refer to a specific task
+ Direct judgment of skill level—criteria define what is being assessed	- Have to infer skill level from total score—criteria do not define what is being assessed
+ Same rubric usable for different tasks (especially useful for portfolios)	- Rubrics must be rewritten for each task
+ Best for "big" student outcomes, e.g., problem solving or critical thinking	+ Good for assessing knowledge and conceptual understanding
- Harder to score	+ Quicker to score
-Harder to get rater agreement, especially at first	+ Easy to get rater agreement
+ Easier to score "non-typical" responses	- Easy to *mis-score* "non-typical" responses
+ Good student instructional tool	- Poor instructional tool

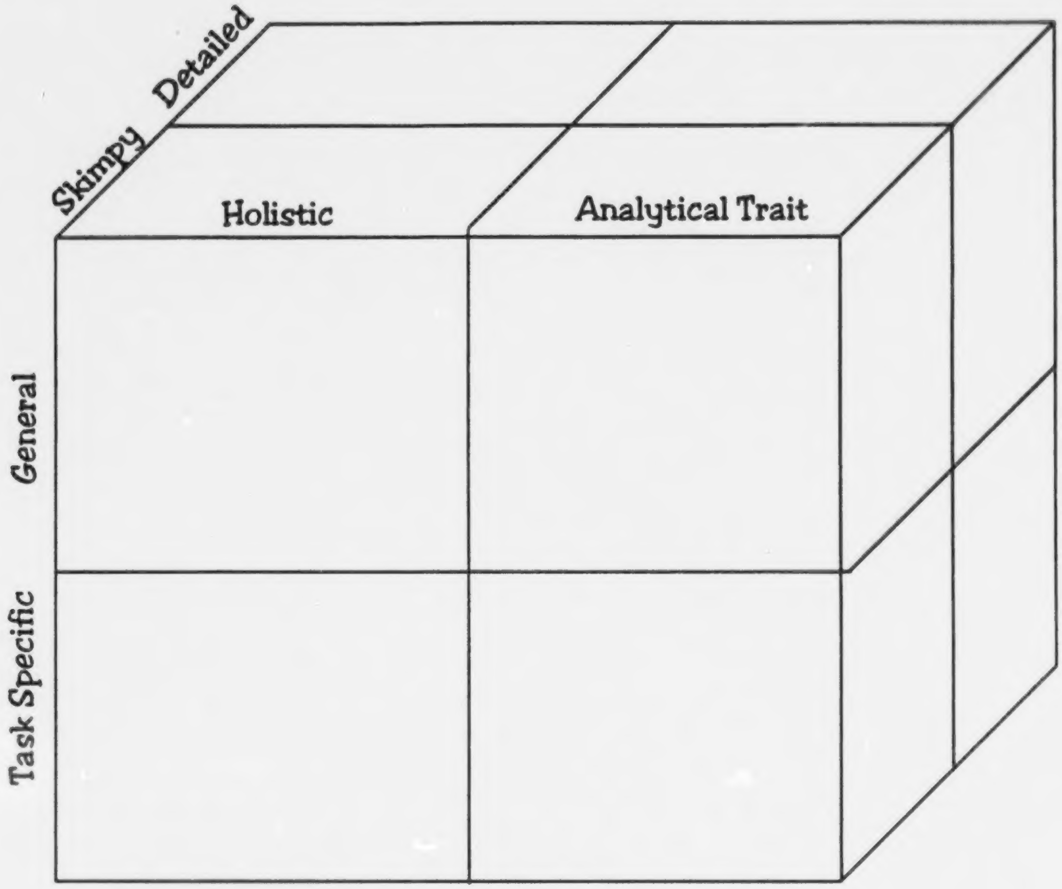
Overhead A1.9,O4

The influencing of instructional practices to date has been served most powerfully by generic rubrics.

Studies of Assessment Reform:
Assessment of Student Performance
OERI, 1997, p. xx

Overhead A1.9,O5

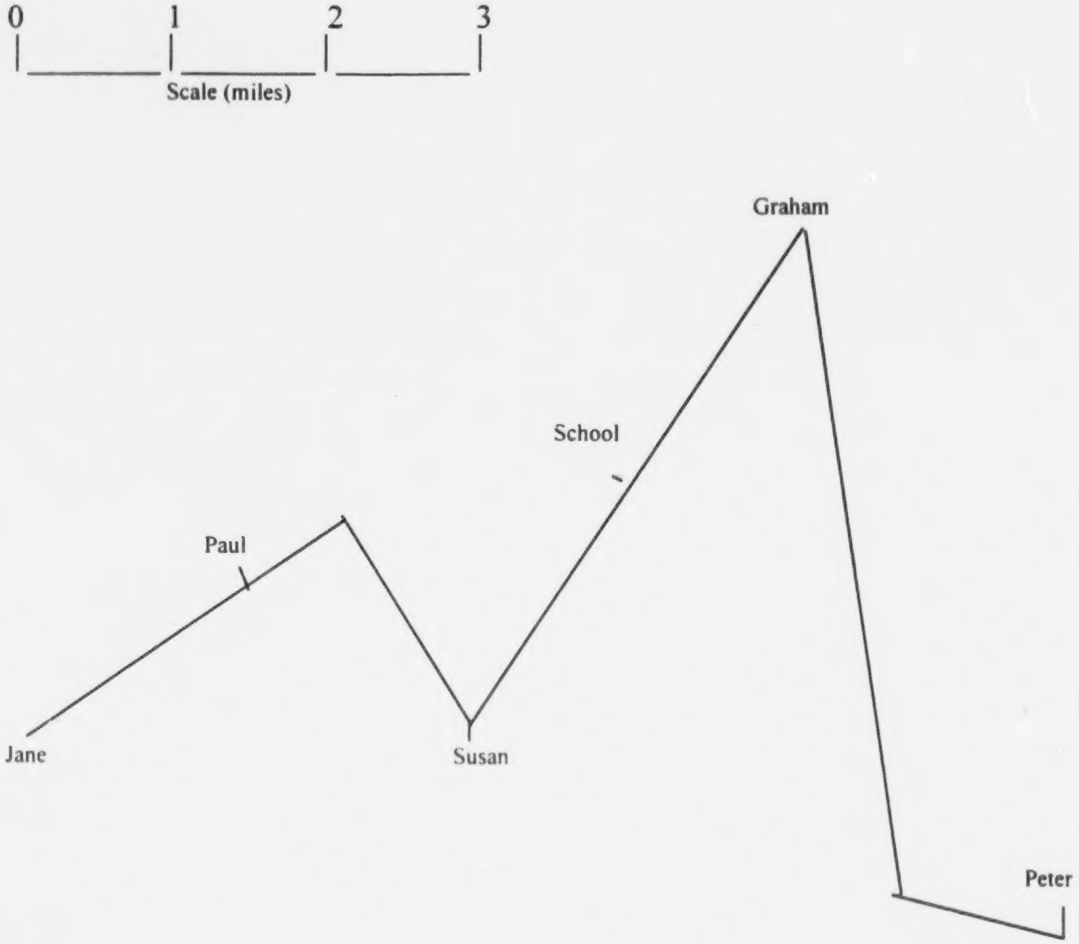
A "Rubrics Cube"


Skimpy
Detailed
Holistic
Analytical Trait
General
Task Specific

Overhead A1.9,O6

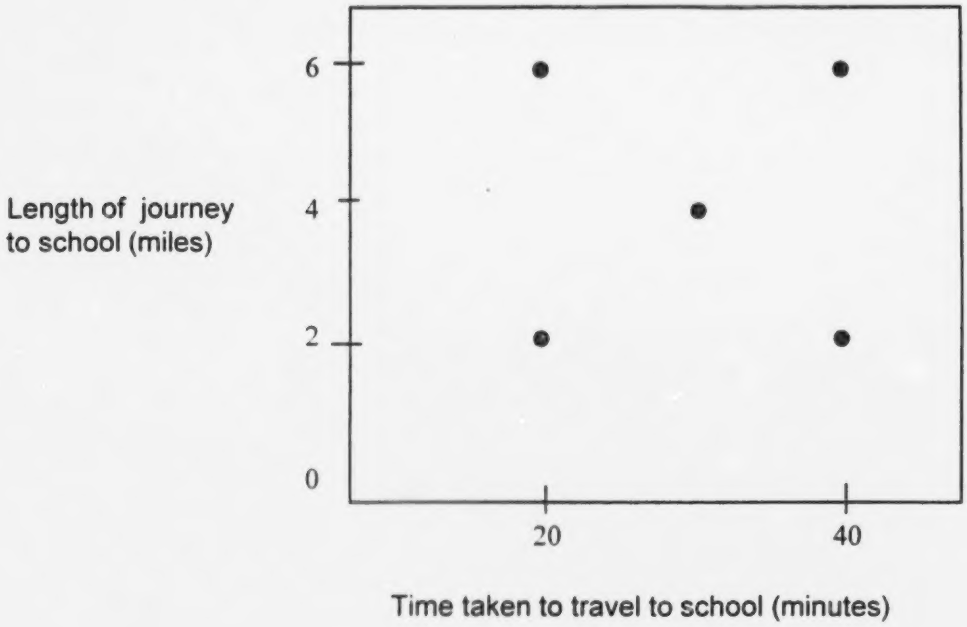
Going To School

Jane, Graham, Susan, Paul and Peter all travel to school along the same country road every morning. Peter goes in his dad's car, Jane rides her bicycle and Susan walks. The other two children vary how they travel from day to day—sometimes walking, sometimes bicycling, and sometimes going in a car. The map shows where each person lives.


0
1
2
3
Scale (miles)
Graham
School
Paul
Jane
Susan
Peter

Handout **A1.9,H1 p. 1**

This graph describes each pupil's journey to school last Monday.

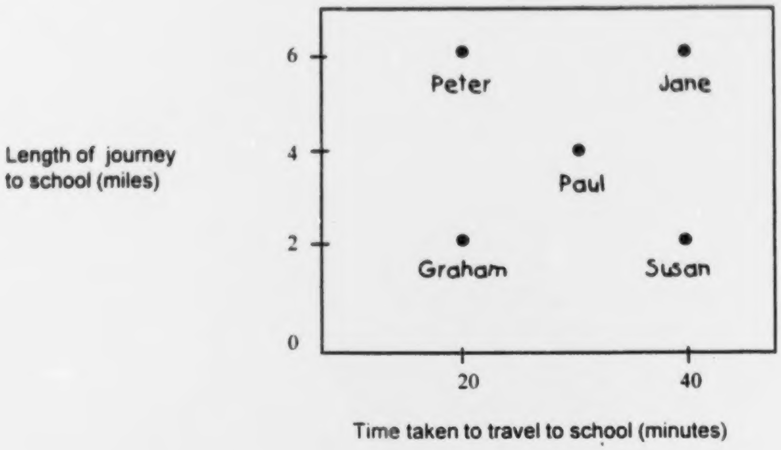


1. Label each point on the graph with the name of the person it represents.
2. How do you think Paul and Graham traveled to school on Monday--walking, bicycle, or car?
3. Describe how you decided on your answer in question 2.

Handout **A1.9,H1, p. 2**

Devon

This graph describes each pupil's journey to school last Monday.



1. Label each point on the graph with the name of the person it represents.
2. How do you think Paul and Graham traveled to school on Monday–walking, bicycle, or car?

I think Graham rode his bike

I think Paul rode his bike

3. Describe how you decided on your answer in question 2.

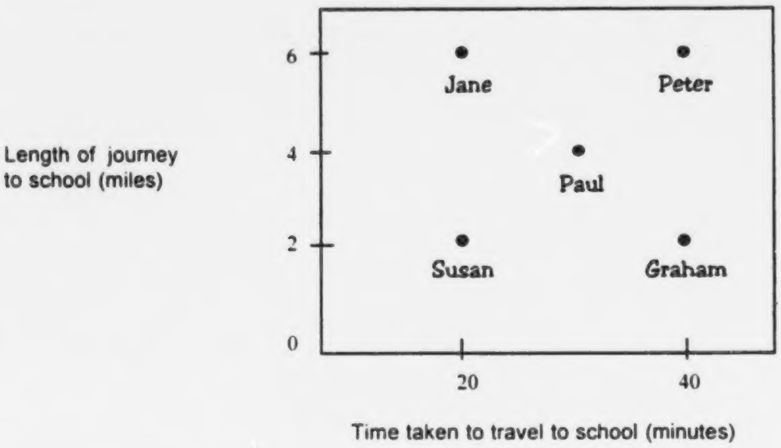
Paul was 4 miles away from school and it took him about 30 min, so it wouldn't be a car because cars are faster and it couldn't be walking be because walking is to slow.

Graham was 2 miles from school and it took him 20 min. to get there. So it couldn't be walking because it is to slow. Driving is to fast, so it must be bike riding.

Handout A1.9,H2, p. 1

Mickie

This graph describes each pupil's journey to school last Monday.



1. Label each point on the graph with the name of the person it represents.
2. How do you think Paul and Graham traveled to school on Monday--walking, bicycle, or car?

Graham rode his bike Paul walked

3. Describe how you decided on your answer in question 2.

Graham lives closer to the school than Peter. And Peter rode in his dad's car. He probly rode his bike. Paul is in the midel and he could probly walk.

Handout A1.9,H2, p.2

236

Task-Specific Scoring Rubric
"Going to School"

1 Point: Shows no understanding of the requirements of the problem. Is unable to label the graph or give a reasonable answer about how Graham and Paul travel to school.

2 Points: Labels one or two points on the graph correctly. Provides an answer to the type of transportation taken by Graham and/or Paul, although it may be incorrect. Provides some sort of rationale for the answer, although it might include faulty reasoning. For example: "Paul lives between Susan and Jane so he probably rides a bike since Jane does," or "Graham lives the same ways away as Susan so he walks."

3 Points: Has three points on the graph labeled correctly. Has identified "bicycle" as the means of transportation for either Graham or Paul. Provides a reasonable rationale for the choice of answer, although it may not be complete. For example, "Graham is faster than Susan so he takes a bike."

4 Points: Has all or most of the graph labeled correctly (4 out of 5). Has identified "bicycle" as the means of transportation for Graham and Paul. Has an explanation for choice of "bicycle" that shows understanding, but requires some interpretation on the part of the rater. For example, "Paul was in the middle and it took him about 30 minutes. Cars are faster and walking is slower."

5 Points: Has labeled the graph correctly. Has identified "bicycle" as the means of transportation for Graham and Paul and is able to explain how the answer was determined. An example of an excellent explanation is: "Susan lives the same distance away as Graham, and Susan walks. If Graham walked it would probably take about the same time. Since it doesn't take as long he probably rides a bike or in a car. Peter takes a car, but goes a lot further in the same amount of time. Therefore, Graham probably doesn't take a car. So, he takes a bike." The response doesn't necessarily have to have this level of detail to be a "5", but the general line of reasoning needs to be there.

Handout A1.9,H3

Task-Specific Scoring Guide #2
"Going to School"
Points for Various Features

Graph: 1 point for each point labeled correctly (5 points possible)

Graham and Paul: 1 point for correctly identifying that each ride their bike (2 points possible)

Rationale: (4 points possible)

4 points: Excellent rationale—includes accurate comparison to other students

3 points: Good rationale—includes comparison to other students, but some parts might be hard to follow or incorrect

2 points: Fair rationale—Provides a rationale, but includes fairly faulty reasoning

1 point: Poor rationale—shows no understanding of the problem

Handout A1.9,H4

Generalized Scoring Rubric
Relative Quality—No Explanations

1 Point: Shows no understanding of the requirements of the problem.

2 Points: Shows a little understanding of the requirements of the problem.

3 Points: Shows partial understanding of the requirements of the problem.

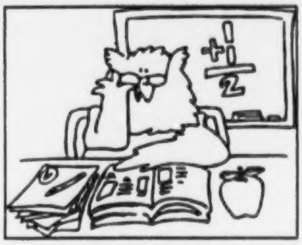
4 Points: Shows considerable understanding of the requirements of the problem.

5 Points: Shows complete understanding of the requirements of the problem.

Handout A1.9,H5

Activity 1.10

Clear Targets and Appropriate Methods—The View From the Classroom



Purposes:

1. To understand what it looks like when teachers have clear and appropriate student learning targets and when targets and purposes for assessment are matched to assessment methods
2. To self-assess one's own current level of understanding of these two important assessment quality ideas

Rationale:

The self-rating forms in this activity are an attempt by *Toolkit98* authors to practice what we preach. Just as we suggest that teachers need to have "clear and appropriate learning targets" for their students, we need to have "clear and appropriate learning targets" for the users of *Toolkit98* that describe what teachers need to know and be able to do with respect to high quality assessment. In other words, if a teacher were "skilled" at classroom assessment, what would he or she know and be able to do? The two rating forms in this activity focus on the skills of "clear and appropriate learning targets" and "aligning targets with assessment methods."

Uses:

This activity builds skills relating to **Chapter 1**, **Key 1**—*Clear Targets* and **Key 3**—*Target-Method Match*. It is best used after reading the chapter text on characteristics of quality assessment or after doing **Activities 1.2**—*Clear Targets, What Types are These?* and **Activity 1.7**—*Target-Method Match*. It is an intermediate level activity.

Materials:

- Overhead projector, screen, transparency pens, chart pack paper and pens

- **Overheads A1.10,O1**—*Developmental Continuums Purposes* and **A1.10,O2**—*Discussion Questions*
- **Handouts A1.10,H1**—*Clear and Appropriate Student Learning Targets Self-Evaluation Form* and/or **A1.10,H2**—*Appropriately Matching Mathods to Targets and Purposes Self-Evaluation Form*

Time Required:

30-45 minutes

Facilitator's Notes:

1. (5 minutes) Use **Overhead A1.10,O1**—*Developmental Continuums Purposes* to go over the goals of this activity as needed.
2. (10 minutes) Ask participants, in small groups, to brainstorm the answers to these questions (see **Overhead A1.10,O2**—*Discussion Questions*):
    a. (For clear and appropriate learning targets): What would a classroom look like that had clear and appropriate learning targets. What would be good answers to the question, "What are your learning goals for students?" What would a teacher know and be able to do if he or she were a master of the targets of instruction?
    b. (For target-method match): What would a teacher know and be able to do if he or she were a master of matching assessment methods to student learning targets?
3. (5-10 minutes) Ask participants to compare their list to **Handout A1.10,H1**—*Clear and Appropriate Student Learning Targets Self-Evaluation Form* and/or **A1.10,H2**—*Appropriately Matching Assessment Method to Targets and Purposes Self-Evaluation Form.*
4. (5 minutes) Ask participants to self-evaluate using the two rating forms.
5. (10-15 minutes) Ask participants to discuss what they learned from this activity.

Clear Targets and Appropriate Methods

Purposes:

1. To understand what it looks like when teachers have clear and appropriate student learning targets and when targets and purposes for assessment are matched to assessment methods
2. To self-assess one's own current level of understanding of these two important assessment quality ideas

Overhead A1.10,O1

Discussion Questions
The View From the Classroom

Clear and appropriate learning targets:

- What would a master of clear and appropriate learning targets for students know and be able to do?
- What would be good answers to the question, "What are your learning goals for students?"

Target-method match:

- What would a master of matching assessment methods to student learning targets know and be able to do?

Overhead A1.10,O2

Clear and Appropriate Student Learning Targets
Self-Evaluation Form

A person who is **skilled** at setting clear and appropriate learning targets knows and can easily articulate the skills and knowledges important for students to learn. Actions typical of such educators include such things as:

- **Knows local definitions** of such terms as **content standards, performance standards,** and **benchmarks**; can provide examples; and can compare classroom/local standards to state/national standards.
- **Can**, without hesitating, **answer the question, "What are your learning goals for students?"** For example, "I have a body of knowledge I want to impart to students, but I also want them to be strong problem solvers, good critical thinkers, have strong communication skills, and be able to work collaboratively."
- **Has *good* learning targets for students**—for example, "I emphasize five types of outcomes with my students: knowledge, reasoning, skills, products, and affect, and here are examples of these in the unit of instruction I am currently teaching." Responses demonstrate a good "mix" of learning targets from knowledge to thinking skills. The educator can state why these are outcomes of enduring importance.
- **Provides thorough descriptive details** about the student knowledge, skill or ability associated with the learning targets. For example, if the learning target is becoming a fluent reader, a person who is skilled would be able to define what a fluent reader knows and is able to do. If the learning target is math problem solving, a skilled educator would be able to define and describe what good math problem solvers know and can do.
- **Explains to students, parents, and others**, in terms they can understand, which learning targets they hope to achieve in their instruction. They can spell it out so that students, parents, and others can accurately explain it back in their own words.
- **Describes the various levels** of student development toward the learning target—for example, the stages of development in becoming a fluent reader or a strong math problem solver. He or she can describe the appropriate level(s) for the grade he or she teaches.
- **Identifies, in student work or actions,** these different levels of development toward the target. For example, a skilled educator knows the specific student words or actions that would indicate their level of reading fluency or their sophistication as a math problem solver. The skilled person can explain how he or she will know if students have accomplished the learning.

Handout A1.10,H1, p. 1

- **Models the skills that students are to master.**

- Without hesitating **shows specifically where during instruction** each learning target is addressed and how a sequence of instructional activities develops the required knowledges and skills over time.

- **Begins with learner targets and plans instruction to achieve them**, rather than planning interesting instructional activities and then going back, if asked, to identify which learning targets were emphasized.

- **Links level of student mastery to an instructional plan** and explains how/why the instructional plan will improve achievement on the target. For example, what instructional experiences would one plan for a student to facilitate, in the most efficient manner possible, improvements in reading fluency or in the ability to solve math problems?

- **Has students who can** use precise and common terminology to describe the quality of their work, articulate the criteria for success, explain what the teacher expects them to learn, and describe how the work they're doing is applicable to their lives.

- **Has posted performance criteria** for things like writing, group skills, and critical thinking.

- **Teaches students** how to use performance criteria to self-assess and improve work.

A person who is **practiced** at having clear and appropriate learning targets knows what it takes to be a master of the targets of instruction, and can do many of things listed above **with some help**. Actions typical of such a person are:

- **Cannot readily define** words like **content standard, performance standard**, and **benchmark**, but can identify examples when the terms are defined for them.

- **Readily lists goals for students**, but sometimes mixes up goals with instructional activities—for example, saying that a goal for students is "literature" rather than saying, "Through the study of literature I want students to know x, y, and z, and be able to do a, b, and c." Or saying that a goal for students is "the research report on the Civil War," rather than saying, "Through doing the research report on the Civil War I want students to know d, e, and f, and be able to do m, n, and o."

- **Tries to explain** to students, parents, and others, which learning targets they hope to achieve in their instruction, but either needs to stop and think about how to best explain it so that students, parents, and others can accurately explain it back in their own words, or sometimes is not entirely successful at the explanation.

- **Links specific learning targets to instructional activities** in general terms, but needs to stop and think about specifically where, in an instructional sequence, different learning targets are emphasized.

- **Plans instruction by identifying activities that would be engaging for students** and then goes back to identify which learning targets were emphasized.

- **Lists features of a quality performance**, but needs assistance to distinguish levels of performance. For example, a practiced educator states that a piece of writing needs to have voice, but needs assistance to articulate what *strong* or *weak* voice looks like.

- **States in broad terms the characteristics of strong or weak performance** (or levels of development), but needs assistance with the specifics and identifying instances and examples of student work or behavior—for example, needs assistance in identifying examples of strong and weak voice in student work, or identifying samples of student work that illustrate strong or weak critical thinking.

- **Uses checklists or other criteria** to judge the quality of student work, but has aspects of the process that sometimes don't work well. For example, consider a teacher who has as a criterion for a paper that it has a beginning, middle and end. Then *any* beginning, middle, and end should get the same credit. We've heard the following, for example: "I put a C on this paper, but the student convinced me it should be an A because after all she did have a beginning, middle, and end. I think it should be a C, but I really don't know why." In this case the teacher is emphasizing the presence or absence of a feature rather than recognizing the need for describing levels of quality of a feature.

A person with an **awareness** of clear and appropriate learning targets for students knows that it is important to be clear on student learning targets, and that it is important to have a good "mix" of learning targets, but **is not sure where to begin or what needs to be done**. He or she would need to be guided to do the following:

- **Clearly state learning targets for students.**

- **Relate goals for students to instructional activities.**

- **Differentiate between goals for students and the instructional activities designed to achieve them.**

- **Describe characteristics of strong performance** on a learning target.

- **Distinguish between levels of quality** (or levels of development), and identify those student words or actions that indicate level.

Handout A1.10,H1,p. 3

- **Describe typical (or ideal) student performance** on a learning target at the grade level the teacher teaches.

- **Develop clear criteria** for student success that go beyond "10 references, at least two pages, and neat."

A person with a **pre-awareness** level of knowledge about clear and appropriate student learning targets does not see the necessity to have clear learning targets, needs to develop a better sense of when a learning target is "clear," or does not have a good 'mix' of learning targets. Several of the following might be true:

- **Counts on the textbook**, or a series of prepackaged instructional materials, to address and define the important learning targets for students. The educator might say, "This is a reputable text book company so they must have a clear idea of which skills they want to develop in students and the best way to develop them."

- **Has students who have difficulty** explaining why they get the grades they do, self-assessing, and describing what quality works looks like (or, who describe quality in terms of superficial features of work, such as: "It has to be two pages," or "I put my name on it").

- **Has difficulty** listing local learning targets for students. When asked about his or her learning targets for students, the person might say things like, "I'm not sure, I'll know as I go along," "What exactly do you mean by outcomes or targets?" "Here are the units I teach," and "Do we really need to pay attention to this stuff?"

- **Lists (or emphasizes in instruction) targets** that are mostly at the knowledge level or that have little long term learning importance.

- **Makes statements like**, "I emphasize these outcomes because I like to teach them," "I emphasize these outcomes because it is what I have always done," or "My goal is to get through the book."

Handout A1.10,H1,p. 4

Appropriately Matching Assessment Method to Targets and Purposes—Self-Evaluation Form

An educator who is **skilled** at matching assessment methods to learning targets and purposes can describe the broad range of assessment options and when to use each. Skilled performance is characterized by:

- **Use of all types of assessment methods** (multiple-choice, matching, short answer, essay, performance assessment, personal communication) in the classroom.
- **Test blueprints** that indicate how each learning target will be assessed and its relative importance.
- **Matching method to purpose and target**. Matching is indicated by such comments as: "I modified my assessment methods because the first time I didn't get what I wanted from students," "The verbs in my outcome statements match the verbs in the assessment approach, e.g., active verbs in an outcome imply performance assessment," "I considered performance assessment, but that wasn't practical, so I compromised by using an essay," and "I don't use short answer conceptual questions because of student lack of experience with this format."
- **Ability to articulate when to use each assessment method**—for example, a person who is skilled can describe a variety of strategies to check for understanding and the advantages and disadvantages of each.
- **Ability to describe each assessment method** and show examples. For example, a skilled person can provide a reasonable definition for "performance assessment" and show some examples.
- **Students who tell you that** the assessment really "got at" what was taught, and the content of the assessment reflects what the teacher said was important.

An educator who is **practiced** at matching assessment methods to learning targets and purposes is trying many different types of assessments, but would like to fine-tune. A proficient person is characterized by one or more of the following:

- **Can pretty much list** various assessment methods, but might have some trouble defining them, or sometimes is inconsistent in identifying examples.
- **Uses various assessment methods**, but has difficulty sometimes explaining why.
- **Sometimes overuses one form of assessment**, say performance assessment or portfolios.

Handout A1.10,H2,p. 1

- **Is somewhat afraid of using "traditional" assessment** methods because of the recent hype about performance assessment.

An educator who is **aware** of the need to match assessment methods to learning targets and purposes has the desire to be more intentional about selecting assessment methods, but doesn't know what needs to be done or where to start. Such a person is characterized by:

- **Has the nagging feeling** that assessment could be done better, but has trouble articulating what to do differently.
- **Wants to try** more performance assessments, but doesn't know how.
- **Doesn't realize** that he or she is already using performance assessments.
- **Notes problems** in prepackaged assessments, but is not real sure what to do about it.

An educator who is at a **pre-awareness** level in matching assessment methods to learning targets and purposes doesn't realize that matching needs to occur. Several of the following might be true:

- **Uses mostly prepackaged assessments** that come with instructional materials without questioning their quality or match to learning targets.
- **Has little variation** in the assessment method used.
- **Justifies assessment methods** by saying things like: "I assess this way because it matches the CTBS," "I like to use just one assessment method so it is consistent all year long," "This is what I've done for the last 25 years," "It's easiest to correct," "If I had better students I could assess differently," "I don't use teacher observation as an assessment tool because it's not on the report card," "I do it this way because the other teacher does it this way," and "The publisher provides this test and it must match instruction because I use the publisher's book."
- **Has students who tell you** that the content of the assessment didn't match what was taught, or didn't emphasize what the teacher said was important to learn.
- **Has assessments that almost always** emphasize knowledge level targets.
- **Has a rigid approach** to assessment

Activity 1.11

Assessment Standards*

Purposes:

1. To review ideas found in the mathematics and science assessment standards and consider how some ideas about assessment have changed
2. To experience an assessment task and relate it to issues surrounding national assessment standards

Uses:

This activity can be used in **Chapter 1** to develop understanding of the national assessment standards or introduce the rationale for changes in assessment. It touches on **Chapter 1**, **Key 1**—*Clear and Appropriate Targets*, **Key 2**—*Clear and Appropriate Purposes*, and **Key 4**—*Bias and Distortion*. The activity might also provide an introduction to characteristics of quality assessment that would be expanded on in **Chapter 3**. This is an intermediate activity.

Rationale:

It is important to consider how assessment has changed as a result of the development of national assessment standards. An understanding of national assessment standards can support decisions that affect student experiences, classroom instruction, educational policy, and the development and use of state and classroom assessments.

Materials:

- Overhead projector, screen, transparency pens, blank transparencies, blank chart paper and pens
- **Overheads A1.11,O1**—*Assessment Standards Purposes*, and **A1.11,O2**—*Possible Focus Group Questions*

- Handouts **A1.11,H1**—*Chickens and Pigs,* **A1.11,H2**—*Plant in a Sealed Glass Jar,* **A1.11,H3**—*Mathematics Assessment Standards*; and **A1.11,H4**—*Science Assessment Standards*

Time Required:

90-120 minutes

Facilitator's Notes:

1. Use **Overhead A1.11,O1**—*Assessment Standards Purposes* to introduce the activity as needed.

2. (5-10 minutes) Divide participants into two sub-groups. These groups need **not** be divided by grade level of students or subjects taught. (There should be one facilitator for each sub-group.) Pass out the **Handouts A1.11,H1-2**, printed double-sided with math on one side and science on the other). One group will do the mathematics assessment task and the other the science assessment task.

   For the mathematics task, allow 5 minutes to work the problem. Instruct participants to attempt to solve the problem in as many different ways as possible. They may work individually, in pairs, or in small groups. Possible answers:

Pigs	14	13	12	11	10	9	8	...	1
Chickens	1	3	5	7	9	11	13	...	27

   For the science task, allow 5 minutes to work and instruct participants to think first about their prediction and then write as thorough an explanation of their prediction as time allows. They may work individually, in pairs, or in small groups.

3. (5-10 minutes) In sub-groups, have participants share three or more different ways of approaching the problem. Chart paper could be used.

4. (5 minutes) As a way to use assessment vocabulary and have everyone start at the same place, have each sub-group brainstorm a list of **facts** (see below) about the assessment they just worked on. Record these **facts** on chart paper. Remind the sub-group that this is not a discussion—listing the facts will be a beginning point for a discussion that follows.

---

[*] Based on work from Betz Frederick, Grand Canyon University, AZ

Facts are statements that can be made about *what is known directly* from the assessment and ***not*** *what can be inferred* about it. Some examples of facts include:

- The assessment asks for an open-ended response.
- The assessment asks students to solve a problem or make a prediction.
- Students are required to write.

5. (5 minutes) As a bridge between the activity and a discussion about the national assessment standards, briefly introduce the national standards (mathematics or science), provide a little background information (see below) about their development, and orient folks to **Handouts A1.11,H3-4**.

Background for Mathematics—The National Council of Teachers of Mathematics (NCTM) has produced three key documents: *Curriculum and Evaluation Standards for School Mathematics (1989), Professional Standards for Teaching Mathematics (1991)* and finally, the *Assessment Standards for School Mathematics (1995).*

Reference Box:

The National Association of Test Directors (NATD) has also established standards for quality assessment that could be used in this activity. For more information, contact Dr. Joe Hansen, Colorado Springs School District No. 11, 1115 N. El Paso St., Colorado Springs, CO 80903, (719) 520-2077, e-mail: jhansen@cssd11.k12.co.us

The Mathematics Assessment Standards were developed because assessment needs to reflect the reform vision of school mathematics. They are based on the premises that:

- All students are capable of learning mathematics, and their learning can be assessed.
- Assessment procedures must no longer be used to deny students the opportunity to learn important mathematics. Instead, assessment should be a means of fostering growth toward high expectations.
- Teachers are in the best position to judge students' progress.

Background for Science—The *National Science Education Standards* (1995) were developed as a contribution to the national goal of achieving science literacy for all students. The *Standards* are directed to all who have interests, concerns, and investments in improving science, education, schools, and society in the United States.

6. (10 minutes) Ask participants to familiarize themselves with the standards by silently reading through the handouts. When finished reading, they can write down some questions or issues they would be interested in exploring with their subgroup.

7. (20-30 minutes) Divide into groups of 4-6 and ask the participants to discuss one or more of the focus questions provided below (**Overhead A1.11,O2**) and/or other questions that they came up with. After the discussion, each small group will be given 3 minutes to share with the entire group of people (mathematics and science). They should prepare a summary of their conversation on chart paper.

Possible focus group questions:

- In what ways does the sample assessment address what is recommended in the national assessment standards? For example, does the assessment promote equity and allow students to reach high levels of accomplishment?
- In what ways does the sample assessment fall short?
- What would you know about students based on their responses to the sample assessment?
- What information could this assessment provide that would be useful in making decisions about what is taught and how it is taught?
- Are there other questions or issues?

8. (20 minutes maximum) Each small group will share a summary of their conversation with the entire large group—3-5 minutes per group. In this way, the science folks will have a chance to hear about some of the issues in the mathematics assessment standards and visa versa. The facilitator can prime the group for the closing activity by asking them to listen for issues discussed by other groups that also applied to their subject's (mathematics or science) assessment standards.

9. (10 minutes) In closing, have participants briefly compare those issues that seemed to be cross-cutting and applied to the assessment standards for both mathematics and science, as well as issues that differed. A note keeper can record these on two pieces of chart paper titled, "Similarities" and "Differences."

Assessment Standards

Purposes:

1. To look at math and science assessment standards and explore how ideas about assessment have changed
2. To relate assessment tasks to issues surrounding national assessment standards

Overhead A1.11.O1

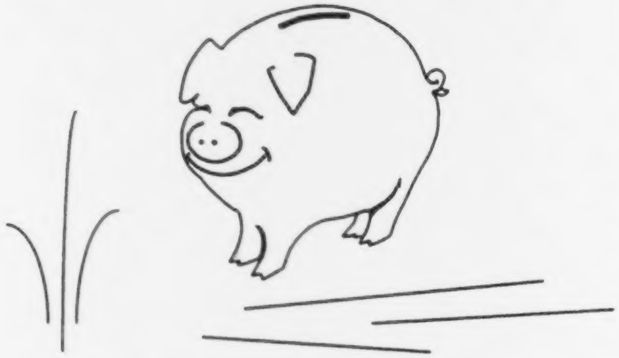
Possible Focus Group Questions

- In what ways does the sample assessment address what is recommended in the national assessment standards? For example, does the assessment promote equity and allow students to reach high levels of accomplishment?
- In what ways does the assessment fall short?
- What would you know about students based on their responses to the sample assessment?
- What information could this assessment provide that would be useful in making decisions about instruction?
- Are there other questions or issues?

Overhead A1.11.O2

Chickens & Pigs

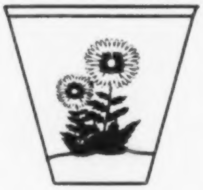
Tom and Sue visited their grandparents' farm. In the barnyard there were some chickens and pigs. Tom said, "I see 18 animals in all." Sue answered, "Yes, and altogether they have 58 legs." How many chickens and how many pigs were in the barnyard? Show your thinking.



Handout A1.11,H1

Plant in a Sealed Glass Jar

Some moist soil is placed inside a clear glass jar. A healthy green plant is planted in the soil. The cover is screwed on tightly. The jar is located in a window where it receives sunlight. Its temperature stays between 60° and 80° F. How long do you predict the plant will live?



Directions:

Use relevant ideas from one or more of the life, physical, and earth sciences to make a prediction. If you are unsure of your prediction, your explanation should state that, and tell what information you would need to make a better prediction. There is no single *right* prediction.

Write your prediction and an explanation supporting your prediction. You may include drawings as part of your explanation.

MATHEMATICS ASSESSMENT STANDARDS—IN BRIEF
National Council of Teachers of Mathematics

The National Council of Teachers of Mathematics (NCTM) has produced three key documents: *Curriculum and Evaluation Standards for School Mathematics* (1989), *Professional Standards for Teaching Mathematics* (1991) and, most recently, *Assessment Standards for School Mathematics* (1995).

The first draft of the *Assessment Standards for School Mathematics* was completed in August 1993. During the 1993–94 school year, critiques and comments were received from more than two thousand reviewers. During summer, fall and winter, 1994-1995, the document was revised and produced.

The Assessment Standards were developed because assessment needs to reflect the reform vision of school mathematics:

- All students are capable of learning mathematics, and their learning can be assessed.
- Assessment should be a means of fostering growth toward high expectations.
- Assessment procedures must no longer be used to deny students the opportunity to learn important mathematics.
- Teachers are in the best position to judge students' progress.

The following key points, made in the *Evaluation Standards*, are expanded upon in the *Assessment Standards:*

- Student assessment should be aligned with and integral to instruction;
- Multiple sources of assessment information should be used;
- Assessment methods must be appropriate for their purposes;
- Instruction and curriculum should be considered equally in judging the quality of a program.

NCTM defines assessment:

ASSESSMENT IS THE PROCESS OF GATHERING EVIDENCE ABOUT A STUDENT'S KNOWLEDGE OF, ABILITY TO USE, AND DISPOSITION TOWARD, MATHEMATICS AND OF MAKING INFERENCES FROM THAT EVIDENCE FOR A VARIETY OF PURPOSES.

If the NCTM's vision is to be realized, all aspects of school mathematics – content, teaching and assessment – need to change on a systemic basis. This includes a shift in the following:

SHIFT IN CONTENT:	*Toward* a rich variety of mathematical topics and problem situations. *Away from* just arithmetic.
SHIFT IN LEARNING:	*Toward* investigating problems. *Away from* memorizing and repeating.
SHIFT IN TEACHING:	*Toward* questioning and listening. *Away from* telling.
SHIFT IN EVALUATION:	*Toward* evidence from several sources judged by teachers. *Away from* a single test judged externally.
SHIFT IN EXPECTATIONS:	*Toward* using concepts and procedures to solve problems. *Away from* just mastering isolated concepts and procedures.

Adapted from *Assessment Standards for School Mathematics.* NCTM 1995, by Betz Frederick.

ASSESSMENT STANDARDS FOR SCHOOL MATHEMATICS

Six ASSESSMENT STANDARDS provide criteria for judging the quality of mathematics assessments. They apply to *all* mathematics assessments.

ASSESSMENT SHOULD REFLECT THE MATHEMATICS THAT ALL STUDENTS NEED TO KNOW AND BE ABLE TO DO.

Assessment should match the current vision of school mathematics and involve activities that are based on significant and correct mathematics. These activities should provide all students with opportunities to formulate problems, reason mathematically, make connections among mathematical ideas, and communicate about mathematics. Skills, procedural knowledge, and factual knowledge are best assessed in the same way they are used, as tools for performing mathematically significant tasks.

ASSESSMENT SHOULD ENHANCE MATHEMATICS LEARNING.

Assessment that enhances mathematics learning becomes a routine part of ongoing classroom activity rather than an interruption. Assessment does not simply mark the end of a learning cycle. Rather, it is an integral part of instruction that encourages and supports further learning. Opportunities for informal assessment occur naturally in every lesson. They include listening to students, observing them, and making sense of what they say and do.

ASSESSMENT SHOULD PROMOTE EQUITY.

In an equitable assessment, each student has an opportunity to demonstrate his or her mathematical power and reach high levels of accomplishments. Because different students show what they know and can do in different ways, assessments should allow for multiple approaches. When students have special needs, provision is made to ensure that they can demonstrate their understanding. Assessment is equitable when students with special needs or talents have access to the same accommodations and modifications they receive in instruction.

ASSESSMENT SHOULD BE AN OPEN PROCESS.

Openness in the process of assessment can be assured in several ways. Firstly, information about the process is made available to those affected by it. Secondly, an open assessment process honors professional involvement and teachers are active participants in all phases. Thirdly, the assessment process is open to scrutiny and modification. Assessments are continually examined for flaws and continually revised to be in harmony with other reforms. Openness in assessment includes informing the public [parents, policy-makers, business and industry leaders, members of the mathematics community, and interested citizens] about the process.

ASSESSMENT SHOULD PROMOTE VALID INFERENCES ABOUT MATHEMATICS LEARNING.

An inference about learning is a conclusion about a student's cognitive processes that cannot be observed directly. The conclusion has to be based instead on the student's performance. Using multiple sources of evidence can improve the validity of the inferences made about students' learning. Mathematics assessment includes evidence from observations, interviews, open-ended tasks, extended problem situations, and portfolios as well as more traditional instruments such as multiple-choice and short-answer tests. The use of multiple sources allows strengths in one source to compensate for weaknesses in others. It also helps teachers judge the consistency of students' mathematical work.

ASSESSMENT SHOULD BE A COHERENT PROCESS.

A coherent mathematics assessment system assures that assessors will develop activities and performance criteria tailored to the purposes of each assessment. Coherence in assessment involves three types of agreement. Firstly, the assessment process forms a coherent whole; the phases fit together. Secondly, the assessment matches the purposes for which it is being done. When the design, evidence-gathering, evidence-interpreting, and action phases of the assessment process are consistent with one another and with the purposes of the assessment, it has educational value. Thirdly, the assessment is aligned with the curriculum and with instruction. Students' learning connects with their assessment experiences.

MAJOR SHIFTS IN ASSESSMENT PRACTICES

TOWARD	AWAY FROM
• Assessing students' full mathematical power	• Assessing only students' knowledge of specific facts and isolated skills
• Comparing students' performance with established criteria	• Comparing students' performance with that of other students
• Giving support to teachers and credence to their informed judgment	• Designing "teacher–proof" assessment systems
• Making the assessment process public, participatory, and dynamic	• Making the assessment process secret, exclusive, and fixed
• Giving students multiple opportunities to demonstrate their full mathematical power	• Restricting students to a single way of demonstrating their mathematical knowledge
• Developing a shared vision of what to assess and how to do it	• Developing assessment by oneself
• Using assessment results to ensure that all students have the opportunity to achieve their potential	• Using assessment to filter and select students out of the opportunities to learn mathematics
• Aligning assessment with curriculum and instruction	• Treating assessment as independent of curriculum or instruction
• Basing inferences on multiple sources of evidence	• Basing inferences on restricted or single sources of evidence
• Viewing students as active participants in the assessment process	• Viewing students as the objects of assessment
• Regarding assessment as continual and recursive	• Regarding assessment as sporadic and conclusive
• Holding all concerned with mathematical learning accountable for assessment result	• Holding only a few accountable for assessment results

Handout A1.11,H3, p. 3

SCIENCE ASSESSMENT STANDARDS—IN BRIEF
National Research Council

Background: The final version of the *National Science Education Standards*, which includes a section on assessment standards, was completed in December 1995.

- The *National Science Education Standards* were developed as a contribution to the national goal of achieving science literacy for all students.
- The *Standards* are directed to all who have interests, concerns, and investments in improving science, education, schools, and society in the United States.

NRC has proposed the following five standards:

A. **Coordination with Intended Purposes:** ***Assessments are consistent with the decisions they are designed to inform.***

- Assessments are deliberately designed. (Science educational data profoundly influence students and those who have responsibility for science education. Therefore, assessments must be well conceptualized and well documented.)
- Assessments have explicitly stated purposes.
- The relationship between the decisions and the data should be clear.
- Assessment procedures are internally consistent. (For example, if the effectiveness of a program is to be measured, then assessments of students' achievement are needed, plus an evaluation of the management system and measurements of teachers' actual increased responsibilities.)

B. **Measuring Student Achievement and Opportunity to Learn:** ***Both achievement and opportunity to learn science must be assessed.***

- Achievement data collected focuses on the science content that is most important for students to learn. (The content parts of the NRC standards portray the aims of science education as rich and varied, encompassing: understanding scientific facts, concepts, principles, laws, and theories; the capacity to reason scientifically; the ability to inquire; the capacity to use scientific principles to make personal decisions and to take positions on societal issues; and the ability to communicate effectively about science.)
- Opportunities-to-learn data collected focus on the most powerful indicators of the students' opportunity to learn. (Such indicators include: teachers' content knowledge and professional skills; the degree to which content, teaching, professional development, and assessment are coordinated; the availability of equipment for student inquiry; and the quality of educational material available.)
- Equal attention must be given to the assessment of opportunity to learn and the assessment of student achievement.

Adapted from the November 1994 DRAFT of *National Science Education Standards*, NRC 1994

Handout A1.11,H4, p. 1

C. **Matching Technical Quality of Data with Consequences:** ***The technical quality of data collected is well matched to the consequences of the decisions and actions taken on the basis of its interpretation.***

- The feature that is claimed to be measured is actually measured.
- Assessment tasks are authentic.
- An individual student's performance is the same on two or more tasks that claim to measure the same aspect of student achievement. (For example, a student's ability to pose appropriate questions is consistent in both physical science and life science given that the tasks are consistent.)
- Students have adequate opportunity to demonstrate their achievements. (This means that the assessment tasks are developmentally appropriate, set in contexts that are familiar to the students, do not require reading skills or vocabulary which are inappropriate to the students' grade level, and are free from bias).
- Assessment tasks and methods of presenting them provide data that are sufficiently stable to lead to the same decisions if used at different times.

D. **Avoiding Bias:** ***Assessment practices must be fair.***

- Assessment tasks must be reviewed for the use of stereotypes, for assumptions that reflect the perspectives or experiences of a particular group, for language that might be offensive to a particular group, and for other features that might distract students from the intended task.
- Large-scale assessments must use statistical techniques to identify differential performance among subgroups that signal potential bias.
- Assessment tasks must be appropriately modified to accommodate the needs of students with physical disabilities, learning disabilities or limited English proficiency.
- Assessment tasks must be set in a variety of contexts, engaging to students with different interests and experiences and must not assume the perspectives or experiences of a particular gender, racial, or ethnic group.

E. **Making Sound Inferences:** ***The inferences made from assessments about student achievement and opportunity to learn must be sound.***

- When making inferences from assessment data about student achievement and opportunities to learn science, explicit reference needs to be made to assumptions on which the inferences are based.

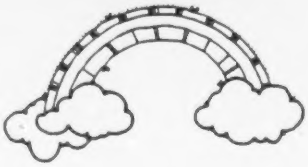
Handout A1.11,H4, p. 2

MAJOR SHIFTS IN SCIENCE ASSESSMENT PRACTICES
(as a result of science education changes in measurement theory and practice)

TOWARD	AWAY FROM
• Using a variety of new methods of data collection, such as performance assessments and portfolios	• Using conventional paper-and-pencil tests almost exclusively
• An understanding that assessment and learning are two sides of the same coin, i.e. students learn from assessment exercises	• A sense that assessment and instruction are independent operations which require being addressed separately
• Greater confidence in assessment procedures that sample an assortment of variables using diverse data-collection methods	• Sampling one variable by a single method
• Assessing all aspects of science achievement—ability to inquire, scientific understanding of the natural world, understanding of the nature and utility of science	• Assessing a narrow portion of science achievement—factual knowledge and application of those facts
• "Authentic assessment" or exercises that closely approximate situations like those encountered in the world outside of the classroom and how scientists work	• Assessments that measure students' achievement in a format that does not mirror situations encountered beyond the classroom, such as using a multiple choice test to assess science process skills
• Greater emphasis on measuring students' opportunities to learn science	• A disproportionate emphasis on assessing student achievement
• Contextualizing student achievement in light of the quality of programs they have experienced	• Interpreting student achievement without considering differences in their experiences
• Assessments that are locally developed and may differ in procedures and instruments—yet they have common meaning and value in terms of the national standards	• Using parallel forms of the same test for national comparisons

Activity 1.12

Assessment Principles: What Beliefs Guide Our Assessment Decisions?



Purposes:

1. To examine underlying beliefs that influence decisions made about assessment
2. To explore equity issues in assessment

Uses:

This activity can be used in **Chapter 1** when a group first begins to examine assessment—as an opportunity to reflect about how previously encountered assessment ideas and information fit, to challenge the beliefs that individuals and groups hold, or as a way to begin a discussion on the rationale for changes in assessment.

Individual principles or clusters of principles can be used as conversation starters for discussion of grading in **Chapter 4**.

Some principles lend themselves readily to discussions of equity in **Chapter 3**.

Rationale:

It's important to surface the varying beliefs about assessment that participants bring to their discussions of alternatives. This activity is intended to help participants look beyond the specific "how to" information and share their thoughts about how assessment should contribute to learning. The set of principles included in this activity was crafted by educators intent on reaching decisions across schools and states. They are not intended to be accepted as is, but to function as a prompt for articulating and challenging beliefs.

Materials:

- Overhead projector, screen, blank transparencies and pens, or chart paper
- **Overhead A1.12,O1**—*Assessment Principles Purposes*, **Overhead A1.12,O2**—*Reflecting on Principles*
- **Handout A1.12,H1**—*Assessment Principles*[1]

Time Required:

20-35 minutes

Facilitator's Notes:

1. (5 minutes) Use **Overhead A1.12,O1**—*Assessment Principles Purposes* to present the purpose of the activity. Make the point that the principles they will be viewing come from one group of educators and may or may not match their own beliefs and experiences. We are not necessarily endorsing them; we present them for discussion.
2. (10 minutes) Pass out **Handout A1.12,H1**—*Assessment Principles* and ask participants to group themselves into sets of 3-4. Use **Overhead A1.12,O2**—*Reflecting on Principles* to prompt participants to look at the match between their beliefs and those expressed in the handout.
3. (10-20 minutes) Invite groups to comment on principles that don't quite jibe with their experience; or that raise questions about current practices. Where you think it's appropriate, add your thoughts about some of the underlying issues that are raised by specific principles.

Variation—Digging Deeper: Assessment Principles and Equity

Time:

30-50 minutes

---

[1] The *Assessment Principles* were developed by Pacific educators in the Federated States of Micronesia (FSM) as part of an FSM-Australia Science Education project titled FASE in 1993. Australian consultants were Jim Hawes and Lee Lee Au Goh of Vital Years.

Uses:

As you use the assessment principles with groups of teachers, parents, administrators, and others who are dedicated to success for all learners, there are some follow-up questions that can focus attention on some of the complexities of equity in assessment.

Facilitator Notes:

1. (10-15 minutes) You may want to begin with the general questions:

   - Are there any principles here that raise equity issues for you?
   - What are your questions or concerns about these principles?

   Ask participants to think about these issues individually and then discuss ideas in small groups.

2. (10-20 minutes) Below are selected principles and related equity questions that come from people in many locations. You can use this list to summarize the small group discussions or to stimulate group thinking. There are far more questions for each principle than you are likely to use in a session; they are possibilities to select among.

Assessment Principle	Equity Question(s)
A2: The main purpose of assessment is to help students learn. When students are assessed well and given feedback about their performance, they find out what they have learned successfully and what they have not. Any weaknesses can then be reduced.	**A2:** Are there differences in how feedback is given and received in the cultures represented in your classroom? What kinds of feedback work (for which students)? • individual? group? positive? critical? • in writing? orally? • by teacher only? by peers? What should feedback focus on? (Errors? Strengths? Movement toward standards?)
A3: Assessment tasks should be designed so that most children in a group do well in most tasks. This takes the threat out of being assessed, and allows children to be motivated	**A3:** Does this principle encourage "dumbing down" in the curriculum? If most students should do well in most tasks, will we avoid really challenging tasks? Will

Assessment Principle	Equity Question(s)
to learn by the regular experience of success and praise.	students with high ability be limited? Is praise always a positive force in student learning? (Does it encourage and motivate all students?
A4: Design/selection of assessment tasks requires a clear idea of the curriculum objectives. Children should only be assessed on knowledge, skills, and attitudes their teachers have given them opportunities to develop, and each task should be well within the capabilities of most students.	**A4:** Does this principle assume that everything of importance is learned only in school? Does it, too, give permission to dumb down tasks for some students? To limit our expectations for them? What are the unique perspectives, ways of knowing and showing learning that students come to school with? How can we build them into our assessments?
A7: The language of assessment must match the language of instruction. If not, then assessment produces unfair and invalid results. Children must be fluent in the language in which they are to be assessed and the level of language used must match their stage of development.	**A7:** What are the language(s) in which our students can best express their learning? How do we assess learners whose language skills in their first language are limited as well as their second language skills? Will it be equitable if we translate? Will it be quality assessment? Will it be equitable if we give students more time? Is that good assessment?
A9: Assessment should focus on each student's achievements, independently of how other students are doing. Constant comparison/competition with classmates can damage the self-esteem and self-confidence of many students.	**A9:** Isn't competition a vital part of success in our nation? If students don't know how they are doing in relation to others, what will be the measure of excellence? (This one can lead into discussion of standards-based assessment.)

3. (10-15 minutes) Invite participants to propose rewordings, deletions or additions to the principles to more fully deal with the essential need for equity in assessment.

Assessment Principles: What Beliefs Guide Assessment Decisions?

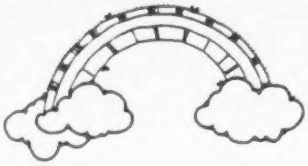
Purposes:

1. To examine underlying beliefs that influence decisions that are made about alternative assessment

2. To explore equity issues in assessment

Overhead A1.12,O1

Reflecting on Principles



1. Which principles affirm your own beliefs?

2. Are there any that challenge your beliefs or experience? What concerns do you have?

3. What are implications of each principle for classroom action and decisionmaking?

Overhead A1.12,O2

Assessment Principles

A1

Learners need to find out often how well they are doing and teachers need to find out how successfully they are teaching. Therefore, regular assessment of student progress and achievement is part of good teaching.

A2

The main purpose of assessment is to help students learn. When students are assessed well and given feedback about their performance, they find out what they have learned successfully and what they have not. Any weaknesses can then be reduced.

A3

Assessment tasks should be designed so that most children in a group do well on most tasks. This takes the threat out of being assessed, and allows children to be motivated to learn by the regular experience of success and praise.

A4

Design/selection of assessment tasks requires a clear idea of the curriculum objectives. Children should only be assessed on knowledge, skills and attitudes their teacher has given them opportunities to develop, and each task should be well within the capabilities of most students.

A5

No single method of assessment can give information about achievement of the full range of learning objectives. Therefore a combination of different methods is vital if we are to get a balanced picture of student development.

A6

Assessment tasks must be presented in a way so that the student is perfectly clear about what is expected, and grades or marks awarded so that the student feels s/he has been fairly treated.

A7

The language of assessment must match the language of instruction. If not, then assessment produces unfair and invalid results. Children must be fluent in the language in which they are to be assessed and the level of language used must match their stage of development.

A8

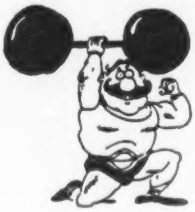
The teacher's unbiased judgments are important in assessment, but students themselves can often be asked to assess their own level of achievement and the achievements of their classmates. They can be (surprisingly) accurate and honest.

A9

Assessment should focus on each student's achievements, independently of how other students are doing. Constant comparison/ competition with classmates can damage the self-esteem and self-confidence of many students.

Handout A1.12,H1

Chapter 2
Integrating Assessment With Instruction



What's in This Chapter?

Have you ever said, "This test sends the wrong messages to students about what is important to know and be able to do," or "This test narrows the curriculum to what's on the test?" If you've ever said such a thing, then you may have experienced a test that does *not* have "consequential validity." This is the technical term for an assessment that has positive influences on teachers, students, and instruction, and avoids negative side effects. Criticisms of past multiple-choice tests, for example include narrowing instruction and reinforcing the notion that everything has a single "right" answer and there is only one right answer. Such tests would not have "consequential validity."

Nothing about alternative assessment automatically ensures "consequential validity." Negative unintended side effects can still occur. Alternative assessments can also do such things as narrow the curriculum if instruction is focused on the small number of performance tasks on the assessment. This chapter considers the ways assessment can influence teachers, instruction, and students. Understanding these influences can help educators build assessment systems that accomplish what is wanted and avoid undesirable side effects—in short assessment systems that are "consequentially valid."

Chapter Goals

1. Present different conceptions of what it means to integrate assessment and instruction
2. Discuss various ways that assessment can influence teachers and students

3. Assist the readers to build a vision of what they would like assessment to accomplish

4. Discuss the assessment design implications of various visions

Chapter Content

A. Readings

What Does It Mean to Integrate Assessment and Instruction?............................Page 4

The first section discusses various ways that alternative assessment can influence curriculum, instruction, teachers, and students. It assists the reader in developing a vision of what he or she would like assessment to accomplish and in understanding why having such a vision for assessment is essential in designing assessment systems.

Integrating Assessment and Instruction: Continuous Monitoring ....Page 10

The second section expands on one of the views in the first section—the notion of assessment being a tool for continuously monitoring student achievement so that instruction can be rationally planned.

Integrating Assessment and Instruction: Using Assessment as a Tool for Learning ....Page 19

The third section expands on another idea voiced in the first section—that alternative assessments offer more than just a way for tracking student progress. If designed properly, they can also be tools for student learning. That is, student achievement can be enhanced by the process of assessing and being assessed.

B. Activities in This Chapter

Activity 2.1 *Sorting Student Work*

This activity is at an intermediate level of difficulty and involves sorting samples of student work into three stacks representing "strong," "medium," and "weak" responses to a performance task. Participants have the opportunity to describe and discuss the characteristics of work that differentiate these stacks. This is an exercise in developing performance criteria, but it also demonstrates how developing performance criteria can help sharpen understanding of the goals held for students, increase teacher expertise, and assist students with understanding how to produce high quality work. The activity includes a description of the steps for developing performance criteria, using self-reflection letters as a running example. The activity uses samples of student work in **Appendix B**, and sample assessments from **Appendix A.** Time: 75-120 minutes

Activity 2.2 *How Can We Know They're Learning?*

This beginning level activity is an adaptation for parents of **Activity 2.1**. Parents have the opportunity to sort student work and distinguish for themselves the characteristics that make it more or less strong. The activity also helps parents understand changes in assessment and why they are occurring. It uses sample assessments from **Appendix A**. Time: Part A, 75-90 minutes; Part B, 30-40 minutes

Activity 2.3 *Ms. Toliver's Mathematics Class*

In this activity, using a video, a teacher demonstrates how she conducts continuous monitoring of student achievement and implements standards-based assessment. It is designed for educators at an intermediate level in their study of assessment. Time: 1½ to 2½ hours

Activity 2.4 *Is Less More?*

Participants experience an integrated, interactive, standards-based mathematics and science lesson and consider the implications of such a lesson for student assessment. It is designed for educators at an intermediate level in their study of assessment. Time: 2 hours

Activity 2.5 *How Knowledge of Performance Criteria Affects Performance*

This activity illustrates the importance of performance criteria in helping students understand the requirements of an assignment or task, and illustrates what can happen when criteria for judging success are not clearly understood by students. It is designed for educators at an intermediate level in their study of assessment. Time: 30-45 minutes

Integrating Assessment With Instruction Readings

What Does It Mean to Integrate Assessment with Instruction?

Integrating assessment and instruction sounds good if one says it fast. But, what does it actually look like day-to-day in the classroom? This section discusses various ways that alternative assessment has the potential to influence instruction, teachers, and students. It's essential to develop a vision of what assessment is to accomplish because the vision informs how assessments will be designed. First we'll look at four possible meanings of "integrating assessment and instruction"—"continuous monitoring," "pushing instruction," "assessment development," and "tools for learning." Then we'll expand on the two of these over which teachers have most control—"continuous monitoring" and "tools for learning."

Please understand, we're *not* saying teachers don't currently integrate assessment and instruction; many teachers have this down to a fine art. We're simply proposing some ideas as a springboard for examining one's practice.

Assessment as "Continuous Monitoring" (View #1)

Assessment can be part of an instructional feedback loop that helps educators monitor student progress toward those goals we've set for them. The monitoring notion applies equally to both large-scale and classroom assessment. At the large-scale level, monitoring is used for making decisions about the quality of the education students are receiving. At the classroom level, teachers need information about the status of student progress toward goals in order to effectively plan the next instructional steps. The teacher who continuously monitors contrasts with the teacher who presents an entire unit and only then assesses, rarely uses assessment results to plan further instruction, and mostly uses assessment results to grade.

Continuous monitoring, itself, is not a new concept. Many teachers already do this very well. What might be new are: (1) the types of student learning targets on which to monitor progress (more emphasis on process skills and the "big" learning targets of collaboration, critical thinking, taking control of learning, etc.); (2) the type of assessments needed for monitoring progress on these kinds of outcomes (more performance-based); and (3) the need, at the classroom level, to collect information more systematically (intuition is no longer enough).

The notion that assessment provides the information that powers instructional decision-making is illustrated nicely in the second section of this chapter. The section describes new thinking about instruction (using science and math as an example) and provides examples of how alternative assessment information fits into the instructional cycle at the classroom level.

Related *Toolkit98* Activities and Chapters

The theme of continuous monitoring is illustrated in **Activity 2.3**—*Ms. Toliver's Mathematics Class* and **Activity 2.4**—*Is Less More?*

Assessment as a Tool to Push Instruction in a Desired Direction (View #2)

This view of integrating assessment and instruction holds that large-scale assessment communicates to teachers what and how to teach. Therefore, if large-scale assessments are designed correctly, instruction will go in the desired direction. In one study done in science classrooms, performance assessment tasks were explicitly designed to model currently known aspects of good instruction—long-term projects, clear outcomes, emphasis on the process of science, and student collaboration. The researchers argued that such modeling is necessary because many teachers don't know how to teach science so that students can attain the learning targets currently proposed for them. The researchers found that teachers changed in their (1) perception of what science is, (2) knowledge of

Reference Box

Brian Gong, R. Venezky, and D. Mioduser, "Instructional Assessments: Lever for Systemic Change in Science Education Classrooms," *Journal of Science Education and Technology, Vol. 1*, 1992, pp. 157-176.

Richard Shavelson and Gail Baxter, "What We've Learned About Assessing Hands-on Science," *Educational Leadership, Vol. 49*(8), 1992, pp. 20-25.

BEST COPY AVAILABLE

science content, (3) awareness of student knowledge needs and attitudes, (4) teaching styles and methods, and (5) professional interaction with colleagues. (Gong, et al. 1992)

A second group of researchers concurs, arguing that current knowledge of how students learn, coupled with the changing nature of learning targets for students, require different instructional techniques, classroom configuration, and assessment. Since teachers tend to "teach to the test," if certain features are incorporated into assessments, teachers will *have* to alter their instruction if only to ensure that their students do well on the test. (Shavelson and Baxter, 1992)

Caution

Building certain features into large-scale assessment tasks may not be enough, however, to change classroom instruction. Based on experiences in Kentucky, Thomas Guskey relates that "teachers did not know how to teach to the performance-based assessments, nor did they believe that they could do so within their current time constraints."

Thomas Guskey, 1996, "What You Assess May Not Be What You Get," in Robert Blum and Judy Arter (Eds.), *Student Performance Assessment in an Era of Restructuring,* Section IV, Article #9. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723, ISBN 0-87120-267-0.

View #2 of integrating assessment and instruction emphasizes the impact of assessment *tasks* on instruction. The idea is that teachers model subsequent instructional activities on the *activities* and *tasks* in the assessment. For example, once teachers see how to incorporate collaboration into an assessment activity, they can apply what they've learned to later instructional activities.

Assessment *Development* as a Strategy to Change Instruction (View #3)

This view is similar to the previous one, except that it is the *development* of the assessment (in addition to the *use* of the assessment) that changes instructional practice. When writing about their efforts to develop meaningful alternative assessments, author after author observes that the process of developing an assessment is at least as useful as having it in place at the end. The following statements are typical:

> *The desired outcome is fully realized only when the participants themselves work through the issues critical to achieving it. The processes of exploration and search* [in assessment development] *are important components in changing educational practice; it is not enough to pass on the wisdom of others' experience.* (Murphy & Smith, 1990)

The dialogue that is happening among teachers is one of the most exciting things happening from this. Portfolio [design] *may actually bring us together more than ever before so we begin to define our criteria much more precisely.* (Sugarman, 1989)

The benefits of portfolios lie as much in the discussions they generate among teachers—and between teachers and students—as in the wealth of information they provide. (Murphy and Smith, 1990)

To our surprise, the [model student papers we chose to illustrate points on the scoring guide] were just one benefit of this entire process. More important, we felt, were the discussion and negotiation that occurred that afternoon. In addition to learning about student writing, we valued the opportunity to talk to one another about the work we do. That sense of being a community of learners continued for the entire year. (Evans, 1993)

🕮 Reference Box

Art Costa and Bena Kallick, *Reassessing Assessment: Seven Issues Facing Renaissance Schools*, in R. Blum & J. Arter (Eds.), *Student Performance Assessment in an Era of Restructuring,* 1996. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

Christine Evans, "When Teachers Look at Student Work," *Educational Leadership 50*(5), 1993, pp. 71-72.

Sandra Murphy and Mary Ann Smith, "Talking About Portfolios," *The Quarterly of the National Writing Project,* October 1990. University of California—Berkeley: Center for the Study of Writing, (510) 642-5570.

Jay Sugarman, "Portfolios Useful Assessment Tool," *Education USA,* November 27, 1989, p. 101.

Developing a meaningful alternative assessment, requires expertise in the content area to be assessed, in how learning occurs, and in assessment. The person who develops the assessment develops the expertise. Why not make teachers these developers, so they can take this increased expertise back to the classroom? Costa and Kallick (1996) are very clear on this point. If meaningful change in assessment is to occur, teachers and students must have the power to assess progress. To accomplish this, schools need to be supported in becoming learning organizations to put teachers back in control of developing assessments.

BEST COPY AVAILABLE

In View #3 of integrating assessment and instruction, developing both the performance *criteria* and *tasks* enhances teaching expertise.

Assessment as a Tool for Learning (View #4)

A fourth view of integrating assessment and instruction sees assessment activities *themselves* as directly influencing students—for better or worse. In this view, assessment is not just an episode of collecting information (as in View #1), nor is it merely a way of influencing future instruction by teachers (as in Views #2 and #3). Rather, assessment, if designed properly, not only monitors student progress toward learning targets, but can actually *help students achieve* the learning targets.

One view of this influence is nicely presented by Stayter and Johnston (1990). They detail the many messages students get from our assessments—what to study, who has control of judging quality, the relative importance of learning facts and applying knowledge, the student role in the classroom, etc. The authors also provide specific examples of how to design assessments to send productive messages. Although their examples come from the area of literacy assessment, their comments apply to other content areas as well.

🕮 Reference Box

Francine Stayter and Peter Johnston, *Evaluating the Teaching and Learning of Literacy,* in T. Shanahan (Ed.), *Reading and Writing Together: New Perspectives for the Classroom,* 1990, Norwood, MA: Christopher-Gordon Publishers, Inc.

Another slant on the direct influence of assessment on students is articulated in the third section in this chapter, which uses mathematics examples to show that clearly articulated performance criteria can promote student learning. In this view, assessments are learning episodes when they help teachers and students clarify the expected outcomes of instruction and promote student self-evaluation and reflection. Thus, it is the *performance criteria* rather than the assessment activities themselves that create learning.

A third way that assessments can directly influence students is illustrated by Gong et al. (1992) who describe how the *tasks* used in their assessment were instructional—students learned from the extended projects that were the basis for the assessment.

In summary, alternative assessments can have a direct impact on student learning (View #4 of integrating assessment and instruction) through the messages they send, through appropriate use of performance criteria that define quality work, and by the instructional value of the assessment task itself.

Implications for Assessment Design and Development

"Hum," the reader might be saying, "this is interesting, but why do I need to know this?" We feel it's imperative to think about one's philosophy and unstated assumptions about "integration" because it will affect how to design and develop alternative assessments. For example, if one would like assessment to *push instruction* in a new direction (View #2), then not only must one enlist teachers in the development, but one must have a clear idea of the specific instructional changes needed so that they can be designed into the assessment.

Likewise, if one wants assessment to function as a *tool for learning* (View #4), one must design certain features into it. For example, one way to increase the instructional usefulness of alternative assessments is to have performance criteria that can be generalized across tasks—they can be used to judge, for example, all student writing and not just the current writing assignment. If we design our scoring procedures to relate only to the current task, students will not be as likely to be able to generalize them to the next task. There is, in fact, research to support the notion that general criteria are most useful instructionally. A 1997 national study of the impact of alternative assessment on teaching and learning concludes that:

Related *Toolkit98* Activities and Chapters

The impact on learning of developing and using performance criteria is illustrated in **Activities 2.1**—*Sorting Student Work*, **2.2**—*How Can We Know They're Learning*?, and **2.5**—*How Knowledge of Performance Criteria Affect Performance.*

Appendix A, Sample A.25—*Acids-Bases* provides an example of an assessment task that is also instructional.

> *The influencing of instructional practices to date has been served most powerfully by generic rubrics. In requiring teachers to design and use assessment tasks that elicit the skills and competency outcomes articulated in the generic scoring rubrics, states, including Kentucky and Vermont have communicated and promoted the educational outcomes valued by the state.*
> (Khattri et al., 1997)

Reference Box

Nidhi Khattri, Alison L. Reeve, Rebecca J. Adamson, 1997, *Studies of Education Reform—Assessment of Student Performance*. United States Office of Educational Research and Improvement (OERI), 1-800-USA-LEARN.

BEST COPY AVAILABLE

Take some time to think about *your* philosophy on integrating assessment and instruction. Do you choose View #1, 2, 3, or 4? A combination? Keeping your intent in mind will make assessment design decisions much easier. The two following sections elaborate on two of the approaches. These, plus **Chapter 3**, will assist in the process of aligning intent with design.

Integrating Assessment and Instruction: Continuous Monitoring

Why Monitor Continuously?

Current understanding of how learning best takes place suggests the importance of constantly monitoring student learning. Students learn best when instruction is:

- generative (guided by student questioning),
- inclusive (attentive to student diversity), and
- constructivist (building on student's prior knowledge).

Good instruction establishes links between new and prior knowledge and supports active, experiential learning. It fosters collaboration and takes into account individual learning styles and stages of cognitive development.

🕮 Reference Box

B. Spector, 1993, "Order Out of Chaos: Restructuring Schooling to Reflect Society's Paradigm Shift," in *School Science and Mathematics, Vol. 93*(1), pp. 9-19.

Assessment must support such instruction. Rather than rewarding only correct responses, for example, assessment must inquire into the reasons behind incorrect responses. Students need multiple forms of assessment for determining learning. And, assessment should stimulate inquiry and foster cooperative learning.

Effective implementation of standards-based education also requires continuous monitoring of student knowledge and skills. If our goal is to make sure all students attain a certain level of competency on well-defined outcomes, it would be foolish to wait for the year-end test to see how successful individual students have been.

Continuous monitoring of student learning means that lines between assessment and instruction fade. The student may be completely unaware of being assessed, instruction modified on the spot, and of further cycles of assessment, instruction, and modification.

Such ongoing assessment activity is designed to create an optimal learning situation for students. It results in continual evaluation and adjustment by the teacher. For example, when students are asked to explain how they arrived at their answer to a problem calling for a comparison of the area of a field to its perimeter, the teacher can assess student understanding of the concepts of area and perimeter, ratio, computation, as well as other skills such as reading comprehension. If a student is not successful in solving the problem because of misconceptions about the meaning of ratio, the teacher can provide experiences that enable the student to master that concept. Conversely, if the teacher determines through this informal assessment that students have already mastered a concept that was planned for a future lesson, he or she can eliminate that instruction.

🕮 Reference Box

For additional detail on "continuous monitoring," see V. Perrone (Ed.), 1991, *Expanding Student Assessment*. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

Here's another example of continuous monitoring using writing assessment. What if students could compare any or all writing to well-defined standards and have feedback instantaneously because they, themselves were doing the assessing? In this situation, students and teachers don't have to wait until the end of the year to see if writing instruction has been successful—performance would be continuously monitored against the standard throughout the year. This is the role of assessment in standards-based classrooms.

In these examples, assessment plays a much larger role than just determining grades. It's a means of planning instruction and tracking student progress toward valued learning targets without having to wait for the unit or year-end "test."

Effective instruction depends upon continuous monitoring. To not do so means continuing to lose those students who have traditionally not performed well—those "at risk."

What Is Integration of Assessment and Instruction When Viewed as Continuous Monitoring?

There are basically two key characteristics of effective continuous monitoring: first, integrated assessment must be a process of inquiry about the reasons for student performance, and second, the assessment must impact the instructional process and, thus, contribute to learning. Following are five instructional situations that provide opportunities for continuous monitoring.

In each of the following examples, key words have been highlighted to indicate approaches that have been successfully employed to integrate assessment and instruction as a continuous monitoring tool: **student-as-teacher, self-assessment, peer assessment, cognitively guided instruction, manipulation of objects to assist in making mental connections, application of new technologies for integrating assessment and instruction,** and **rubrics/criteria**.

Example 1
Excerpt from "Learning Mathematics with Understanding"

A roadblock in the sequence of development of mature problem solutions is the use of symbols. Consider how Ms. Miller helped her children to develop their own understanding of the complex process of writing number sentences.

The children in Ms. Miller's class could solve "joint result-unknown" problems and "separate result-unknown" problems quite well. Some children used direct modeling or counting on, and a few, derived facts. Some children recalled facts with small numbers. No one used written symbols. Ms. Miller asked a small group to join her near the blackboard and asked them to solve a "joint result-unknown." Ms. Miller then showed the children how that problem looked when represented by symbols and asked the children to explain the symbols. One of the children gave the following explanation for transferring what she did with her counters to the number sentence. "Numbers show the same things the counters show." Ms. Miller continued to have the children solve problems and to write the number sentences on the board. She always asked the children to relate their solution to the number sentence.

Ms. Miller decided that this group was ready to move on to using symbols. For the next several months, she carefully monitored their understanding of the relationships between strategies such as direct modeling and the writing of symbols. She never expected children to perform operations with symbols that they could not perform readily by direct modeling or counting strategies. She continually pointed out the relationship of symbols to their less mature strategies. She believed that the children would eventually construct the relationship between symbols and other strategies and thus come to use the symbols with understanding.

Fennema, Carpenter, and Peterson, 1989. Pp. 195-225
(permission to use granted by Fennema)

Note in this example that the assessment is a continual process aimed at identifying the students' cognitive level so that instruction can be modified accordingly. Without the continuous assessment, the instructor would not know whether students were cognitively ready to move on to the next stage of understanding. The **cognitively guided instruction model** has both of the key characteristics of integrated assessment and instruction—it is a circular process of inquiry used to modify instruction.

🕮 Reference Box

E. Fennema, T. Carpenter, and P. Peterson, 1989, "Learning Mathematics With Understanding: Cognitively Guided Instruction," *Advances in Research on Teaching, Vol. 1.* JAI Press Inc., (203) 661-7602.

Measuring Up: Prototypes for Mathematics Assessment, 1993. National Academy Press, (800) 624-6242.

Example 2
Senior Physics

In one **student-as-teacher** integrated assessment/instruction task developed by a high school teacher, senior physics students used an inflatable portable planetarium to learn and teach basic astronomy facts. In pairs, they researched and prepared an astronomy teaching presentation for one or more science classes in the school. A teacher-developed test was used to determine whether students had enough astronomy content knowledge to feel comfortable teaching basic astronomy facts to others. If not, basic concepts were re-taught. Using various teacher-developed **rubrics**, students were also continually assessed on their research and collaborative skills. During planetarium presentations, students were further assessed on their responses to questions from the audience and adaptations of their presentation to the audience—skills particularly enhanced by content knowledge and understanding.

Example 3
Chemistry Demonstrations

In another **student-as-teacher** approach that also combined **self- and peer-assessment**, a team of high school science teachers developed an integrated assessment and instruction activity in which pairs of students prepared chemistry demonstrations focusing on selected biological and chemical processes. They then performed the demonstrations for younger students in other science classes, responding to their questions, while class members videotaped the productions. Class members subsequently critiqued their own and each other's performances using the videotapes and **criteria** developed jointly by the students and teacher.

Example 4
Bridges

The Mathematical Sciences Education Board provides prototypes of integrated instructional/assessment activities that support the NCTM Standards. In the prototype "Bridges," elementary students are provided with **manipulatives** and student worksheets with the following instructions:

You are an engineer who wants to build different kinds of bridges. The bridges will be made of colored rods. The first bridge you are to build is a 1-span bridge made with one yellow rod and two red rods... The yellow rod is called a span and the red rods are called supports. Since the yellow rod is 5 cm long, the length of the bridge is 5 cm... As you build bridges in the following activities, think of a way to keep track of the number of rods of different colors you use. Your goal is to find out how many rods of each color you would need to build a bridge of any size. ("Measuring Up," 1993, p. 44)

Students then proceed to work through the activity. They are encouraged to monitor their own understanding and to "decide for themselves what they need to do to understand the situation" (p. 49). The task, while focusing on "the generation and identification of patterns in a numerical setting...is designed so that a variety of approaches to the computations are feasible.... More important than the computation itself, however, are the students' decisions about what computations are appropriate, and the connection between the computation and the dimensions of the bridge" (p. 49). It is these ongoing decisions that must be monitored with ongoing assessment.

Example 5
Fox Rot

The advent of new technologies has the potential to further expand our ability to integrate assessment and instruction. An example of a prototype instrument currently under development is the *Multimedia-Based Assessment Sampler* which uses a computer linked to a video disc player to simultaneously instruct and assess student process skills. In the task "The Fox Rot" (see **Figures 2.1, 2.2**, and **2.3**), students observe photos of a fox decomposing, place them in chronological order, and explain reasons for their ordering. Students may then view a short movie sequence as a form of self-evaluation, after which they can modify their responses. Responses may either be elicited verbally by the teacher, or they can be entered into a CLARIS Works program for printing or electronic network access by the teacher. Thus, assessment is by the student and/or teacher, with immediate feedback and modification of instruction readily administered.

Figure 2.1
Initial Prompt and Rubric–The Fox Rot

Performance Task—The Fox Rot

Part 1: Field Research

PROMPT: Observe the 4 photos of a fox decomposing. Place the photos in order from start of decomposition (top) to end of decomposition (bottom). List the observations that you make to support the order in which you place each photo. Be sure to compare data from each photo.

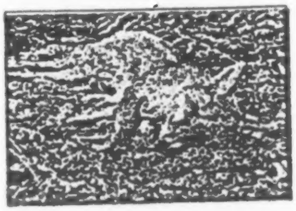
RUBRIC—Observing/Inferring:
TX Essential Element 2 (B, C, D)—Acquire data through senses.
TX Essential Element 4 (B, D)—Communicate data and information.
TX Essential Element 6 (B)—Draw logical inferences.

POINTS	**CHARACTERISTICS**
4	Observes/infers properties of the fox (which may include its surrounding environment) for all four photos.
3	Observes/infers properties of the fox (which may include its surrounding environment) for three photos.
2	Observes/infers properties of the fox (which may include its surrounding environment) for two photos.
1	Observes/infers properties of the fox (which may include its surrounding environment) for one photo.
0	Makes no observations/inferences.

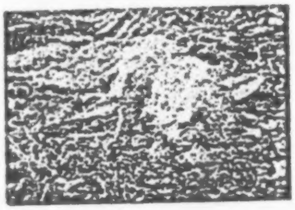
–Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Figure 2.2
Sample Student Response

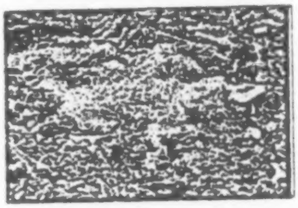
My Fox Story



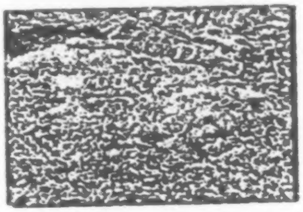
I think that this is the beginning of the fox rotting. It pretty much looks like the fox might be alive but just sleeping. There are no signs of it decomposing yet.



Here you can see how the fox is changing. It is kind of puffed up. I've seen dead animals in the neighborhood and on the side of the road that are all swollen after being dead awhile...especially if its in the summer. All the green plants around the fox make me think it is summer.



Really gross!!! The fox's body is covered in maggots. They must be eating it because a lot of the fox is missing. Or, I guess some scavenger might have come along and eaten part of it before the maggots got there.



Its almost all gone now. There's lots more maggots too. If you didn't know it was a fox at the start you almost couldn't tell now.

The End.

Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Figure 2.3
Verification Prompt and Rubrics

Part 2: Verification

PROMPT: View the time lapse sequence of a fox decomposing. You may view it as often as you like. Compare the sequence of photos you made in Part 1 with the time lapse sequence. Does your photo order match the time lapse sequence? Explain your answer using observations to support your conclusion.

RUBRIC—Conclusions:
TX Essential Element 4 (F)—Interpret the arrangement of data on…visuals.
TX Essential Element 6 (D)—Draw conclusions from the observed data.

POINTS	CHARACTERISTICS
4	Draws conclusion. Bases conclusion on observations comparing all still photos of the fox with the time lapse segment.
3	Draws conclusion. Bases conclusion on observations comparing three still photos of the fox with the time lapse segment.
2	Draws conclusion. Bases conclusion on observations comparing two still photos of the fox with the time lapse segment.
1	Draws conclusion. Bases conclusion on observations comparing one still photos of the fox with the time lapse segment.
0	Draws no conclusion.

RUBRIC—Sequencing:
TX Essential Elements 3 (A)—Classify, order, and sequence data.

POINTS	CHARACTERISTICS
3	Sequences all 4 photos to match actual sequence.
2	Sequences 2 of 4 photos to match actual sequence.
1	Sequences 1 of 4 photos to match actual sequence.
0	Sequences none of 4 photos to match actual sequence.

Courtesy of AOP Hub/Midwestern State University Assessment Consortium. "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

The approaches highlighted in the above examples by no means constitute a comprehensive list of instructional and assessment techniques that incorporate the two essential components of continuous monitoring. Other approaches include:

- Observation of students during regular instructional activities
- Verbal interactions with students—interviews, informal oral questioning, oral examinations
- Written text records—essays, stories, descriptions of processes
- Drawings and illustrations—diagrams, conceptual maps and webs, graphs, sketches, detailed drawings
- Demonstration of skills—dramatic responses, successful experiments, use of manipulatives

Any of these tasks can serve, either singly or in combination, as vehicles for continuous monitoring. They need only be applied with the purpose of inquiring about the status and/or process of student learning, with the intention of modifying instruction based on that assessment, in a continuous fashion.

Some researchers and educators look forward to the day when there will be a "seamless web" of instruction and assessment in which assessment is no longer a distinct activity but is built into students' regular learning experiences, virtually indistinguishable from instruction. In this vision, students and teachers receive ongoing feedback as needed during the instructional process. Assessment for classroom purposes (as opposed to external monitoring purposes) is a routine, non-threatening process. This vision calls upon students, teachers, and administrators to look at assessment in a different way. Teachers must

🕮 Reference Box

M. Land and T. Peters (in progress). *Multimedia Based Assessment Sampler*, AOP Hub/Midwestern State University Assessment Consortium, Midwestern State University, Division of Education, Wichita Falls, Texas, (800) 842-1922.

N. Kober, 1993. *EdTalk. What We Know About Science Teaching and Learning.* National Education Knowledge Industry Association—NEKIA, (formerly Council for Educational Development and Research), (202) 429-5101.

Caution

Although any of these methods *could* be used to continuously monitor student achievement, at least two things must be true to be successful. First, the method must be consciously chosen to match the skill being assessed, as stressed in **Chapter 1**. Second, the teacher and students need to have a clear idea of what to look for in the work—in short, criteria—to judge when understanding has occurred, progress made, or quality demonstrated. Tasks such as those described in this section are, by themselves, not enough to continuously monitor progress.

BEST COPY AVAILABLE

> *understand that the primary information will not be quantitative, but will provide a rich portrait of student strengths and weaknesses. In addition, assessment to inform instruction does not have to compare students to one another, which means that student approaches and responses to a problem may look very different and still be correct...[Such assessment] should measure the processes students use as well as the answers they reach. It should measure all of the goals of the curriculum, not just a few. It should address both group activities and individual ones. It should be developed by, or with ample input from, teachers and should include teacher professional observation and judgment. Perhaps most importantly, it should draw upon information from multiple assessment sources, including but certainly not limited to tests.... Finally, it should have a strong self-evaluation component for both students and teachers.* (Kober, 1993, p. 62)

⊂ Related *Toolkit98* Activities and Chapters

The theme of continuous monitoring is illustrated in **Activity 2.3**—*Ms. Toliver's Mathematics Class* and **Activity 2.4**—*Is Less More?*

Integrating Assessment and Instruction: Using Assessment as a Tool for Learning

Can There Really Be Assessments That Promote Learning?

Mention the word "assessment" to most students and teachers and they will immediately respond with the concept of "tests"—the kind used to monitor and track student performance. *TESTS*—the type given to produce grades, select students for special programs, and report to the Board of Education. Wouldn't it be exciting, however, if the word "assessment" was associated not with traditional testing, but with a different process—one that not only provides an opportunity for teachers and students to thoughtfully examine what has been learned, how we know it has been learned, and what needs to happen next to reach a new level of learning, but also is an episode of learning in and of itself?

The following performance assessment in mathematics is an example of such an assessment. It illustrates how assessment and instruction can be integrated as a "tool for learning."

Performance Assessment: Mathematics

Performance assessment is "assessment based on observation and judgment" (Stiggins, 1994); we observe a behavior or examine a work product and judge its quality. This is frequently done in the classroom by teachers, and is a common occurrence in daily life as well—just think of the Olympics, job performance reviews, and the driver's test. These kinds of judgments can be very subjective. The goal is to make these essentially subjective judgments as objective as possible. This is done through the development of performance criteria that clearly define characteristics and indicators of quality student work.

The criteria developed to judge the quality of student work are also critical to making performance assessment a powerful tool for learning because the criteria clearly communicate what is valued in the work—what it takes to be successful. These criteria for quality are teachable to students, allowing them access to the "secrets" of success. (In fact, students can be full partners in the development of the criteria for success.)

Thoughtfully written and teachable criteria support standards-based teaching. When students and teachers begin to share the same vision of success, they quickly develop a vocabulary for discussing this vision. All are working toward the same clear and identifiable standard. This improves student achievement.

To see how this works, we'll use a mathematics assessment example. Let's begin with analyzing samples of student work. Look at the two student solutions in **Figure 2.4** and take a moment to jot down your reactions to them. What did you notice? Are there any obvious strengths? Weaknesses? What features did you respond to? Which solution did you think was the best? Why?

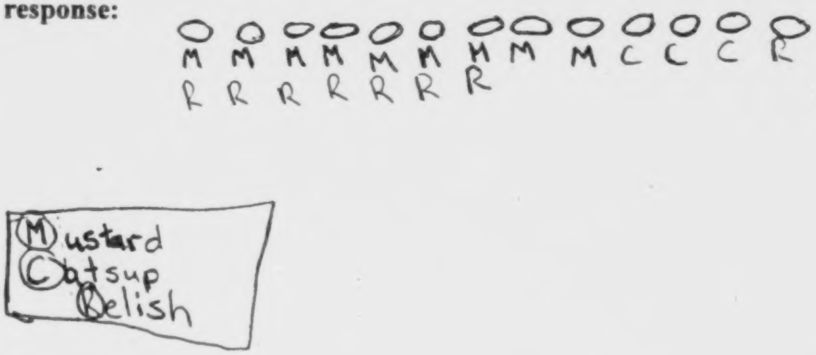
🕮 Reference Box

Richard Stiggins, *Student Centered Classroom Assessment, 2nd Ed.*, 1997, Columbus, OH: Macmillan.

Figure 2.4
Mathematics Applications—Grade 4

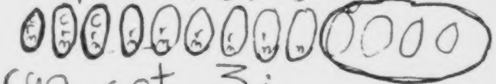
Problem: A family of six orders 13 hot dogs. Nine hot dogs have mustard, three have catsup, eight have relish. Explain to the clerk how many hot dogs have no mustard, relish or catsup. You may use drawings in your explanation.

First student response:

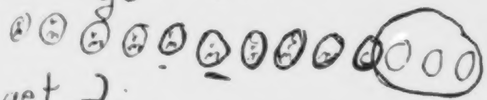

M M M M M M M M M C C C R
R R R R R R R
Mustard
Catsup
Relish

Second student response:

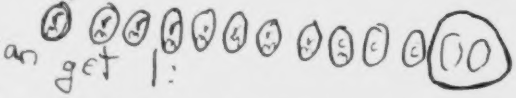
"I would like 13 hot dogs. 9 with mustard 3 with catsup. and 8 with relish." It would be possible to have 4 plain hot dogs or 3 or 2 or 1 or none. This is how you can get 4 plain hot dogs:



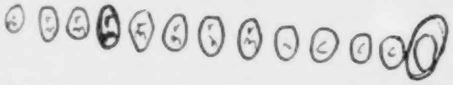
This is how you can get 3:



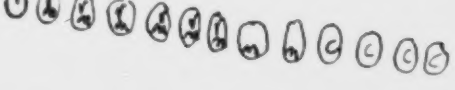
This is how you can get 2:



This is how you can get 1:



You can get 0 this way:



It depends

Most teachers and students examining solutions such as these tend to like and dislike the same sorts of things. Typical comments for the first solution include: *vague, may not have understood the problem fully*, and *had to read a lot into the solution in order to understand what the student did.* Typical comments for the second solution include: *understood the problem fully, clear solution, communicated well in pictures and words, logical, detailed, organized, insightful, correct answer, explained thinking, used all relevant information, distinguished relevant information from irrelevant information, used data accurately, correct process, on topic, creative, focused, showed process, used pictures correctly, good explanation*, and *went beyond the minimum.* How did your list of comments compare to these?

Comparing solutions to mathematics problems helps us all sort through the complexity of features that make solutions successful. Such exercises can help teachers and students explore what makes a solution work. In fact, the same sets of features recur on brainstormed lists from teachers across the country. They form the basis for performance criteria for mathematics applications problems in many places. For example, **Sample A.18**—*Mathematics Problem Solving* (in the *Sampler* **in Appendix A**) shows a "four-trait" model for assessing such problems. There are four traits because teachers in Oregon felt that their list of comments tended to fall into four major areas: conceptual understanding, procedural knowledge, problem-solving strategies, and communication. Look at the features other teachers tend to place under each trait **(Figure 2.5)**. Do the comments you made on the work in **Figure 2.4** fit in the same way?

⊂ Related *Toolkit98* Activities and Chapters

Appendix A—*Sampler* contains several other rubrics/criteria for judging solutions to math problems. Compare them. Do the same sets of features tend to recur?

Activities 2.1—*Sorting Student Work*, **2.2**—*How Can We Know They're Learning?* involve sorting student work, generating lists of desirable features, and beginning the process of developing performance criteria.

The four-trait model for analyzing mathematics applications is, thus, basically an attempt to systematically write down *what we, as educators, already value* so we can be consistent in our vision of success across students, across assignments, and across time. And, equally importantly, so students can share in the vision of success and not have it constantly shifting in unknown ways between teachers or assignments.

Figure 2.5
What Teachers of Mathematics Value

Conceptual Understanding

- Used all relevant information
- Distinguished irrelevant information
- Understood the problem
- Creative
- Used data accurately

Procedural Knowledge

- Used pictures correctly
- Correct answer

Problem-Solving Strategies

- Logical
- Correct process
- On-topic
- Insightful
- Beyond the minimum
- Focused

Communication

- Detailed
- Organized
- Good explanation
- Communicated well in pictures and words
- Clear solution
- Showed process
- Explained thinking

Value of Criteria to Teachers. Have you ever been nervous grading students' open-ended mathematics problems? Those who have rated or scored mathematics during a direct mathematics assessment using criteria such as the four traits unanimously agree that the very process of assessing teaches a great deal about what makes mathematics solutions successful. Raters have to internalize a set of criteria so thoroughly that they can be consistent both with themselves over time and with other raters. During the process of scoring hundreds of papers, raters get to see lots of examples of what good and poor mathematics actually looks like and to analyze it systematically to determine why it is good or poor. These teachers return to the classroom much more confident in commenting on student work and helping students understand what it takes to produce sound, efficient, and understandable solutions to mathematics problems.

Great! As teachers become more consistent in their analysis of student math work, they gain the confidence to share their insights with students. If assessing brings about a shared vocabulary and a deeper sense of what it takes to solve mathematics problems well, then wouldn't students also benefit from the same process? The answer is YES.

Value of Criteria to Students. Students can also learn to analyze sample work and develop a systematic vocabulary for describing quality so they can develop their own sense of what it means to present a good solution. Consider the following quotes from the area of writing assessment and think of how they might also apply to mathematics problem solving, science process skills, group collaboration skills, and other complex goals we have for students.

> *I want (students) to see evaluation in its best sense—a source to inform teaching and learning. To that end we develop a vocabulary for commenting on the admirable and problematic aspects of writing....The more we examine samples, the richer and more helpful this language of evaluation becomes.* (Erikson, 1992)

> *Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses.* (Lucas, 1992)

> *Scales, criteria, and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them.* (Hillocks, 1986)

The goal of assessment should not be only to earn grades. The goal of assessment should be to help students know "how to win again." Clear performance criteria have this effect.

This is instructional assessment. Students learn something in the process of assessing and being assessed. They learn about mathematics problem solving and how to apply this knowledge to their own and others' works. They learn what effective mathematics solutions look like and what to do to make their own mathematics skills stronger.

Reference Box

G. Hillocks, Jr., 1986. *Research on Written Composition.* National Council of Teachers of English, (800) 369-6283.

C. Lucas, 1992. "Introduction: Writing Portfolios—Changes and Challenges," in K. Yancey (Ed.). *Portfolios in the Writing Classroom*, pp. 1-11. National Council of Teachers of Mathematics, (703) 620-9840.

M. Erikson, 1992. "Developing Student Confidence to Evaluate Writing," in *Quarterly of the National Writing Project & The Center for the Study of Writing and Literacy, Vol.14*, pp. 7-9.

Beyond Mathematics Assessment

The point here is that good performance criteria can be more than just a tool for teachers (or districts or states) to monitor student ability to write, problem solve, be good collaborators, or be critical thinkers. They also help both teachers and students conceptualize and define standards, especially in hard-to-define areas such as critical thinking, problem solving, collaborative learning, and communication. They provide a vocabulary for discussing features of work or behavior that have, in the past, been subtle and ill-defined.

Good performance criteria help teachers answer such questions as: *What is expected? What are our standards? What does good performance look like? What do I want to accomplish? What kind of feedback do I give to improve student work next time?* They help students to answer similar questions: *What is expected? What does good performance look like? How will I know when I'm successful? Why did I win? How can I win again?*

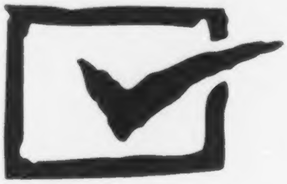
There are three essential features to make it work: (1) having a clear picture in our own heads of the criteria for success, (2) making sure the criteria are right—they really do describe the important features of quality, and (3) letting students "in on it."

1. **Having a clear picture in our own heads**. Criteria are, in effect, the final and most complete definition of district and teacher goals for students. Notice that it is the **performance criteria** and not the **performance task** that have this function. It is the performance criteria that define what we value in student work, define our targets and standards, and are useful in determining when we meet standards.

   In many discussions of integrating assessment and instruction, criteria are left out. There is a lot of talk about the tasks that should be given to students and how one must watch performance on these tasks to monitor where students are. But the question is: *"What, exactly, do you watch for?"* to determine progress, know when students are ready to move on to the next topic, or determine when mastery has occurred? Criteria tell us what to look for. Overt and public criteria—arrived at through discussion and collaborative learning and accessible to teachers, students, and parents—bring judgmental assessment from the realm of individualistic impressions to the realm of systematic and generalizable documentation of learning. They also result in assessment that promotes learning.

How does one know when criteria are clear enough?

- Everyone can list the criteria
- Everyone gives the same definitions for features listed in the criteria
- Everyone can identify the same samples of strong and weak student performance
- Everyone can use the same words for describing why the samples are strong or weak



2. **Making sure the criteria are right**. Is handwriting a telling and important feature of math problem solving? Is neatness the most telling and important feature of writing? Do all the sentences in writing have to be complete, or is sentence quality more determined by the flow of the sentences than whether they are complete or sentence fragments? Should setting up the problem correctly be as important as calculating the right answer? A two-minute speech might be desirable, but it that the most telling feature of a quality oral presentation?

These are the types of decisions that have to be made when developing criteria. The criteria have to describe what everyone really means by quality or they won't help anybody. If a teacher or student believes that an essential component of quality has been *left out* or that an unimportant component of quality is prominently *in* the criteria, there will be no buy-in. If criteria leave out an important feature of quality, like voice in writing, teachers might not teach voice. What they see is what *you'll* get, so criteria have to embody the real indicators of a sound performance and leave out that which is not important.

How does one know when criteria are the right ones?

- No performance that is weak has to be judged strong using the criteria.
- No performance that is strong has to be judged weak using the criteria.
- After repeated use, one can't think of anything else that needs to be included or left out.
- The research literature supports the features included.
- Teachers, students, and parents consistently reproduce the criteria when asked to sort work for quality and describe the reasons for their judgments.

3. **Letting students in on it—or—students must be involved in assessing**. Students learn to the extent that they are the ones doing the judging.

> *By encouraging students to engage in self- and peer evaluation teachers empower students to take control of their own learning. When students help determine the criteria for assessment they can make reasonable decisions about the quality of their own work. By engaging in self-assessments, students learn they are ultimately responsible for their own learning. It is in this area that portfolios are so powerful.* (Tierney, et al. 1991, p.59)

> *To be successful, students must learn how to evaluate rather than how to be evaluated....Evaluation is routine, something to be done at every step of the way as they examine decisions, observations, data, ideas and outcomes....[When] evaluation is a key process, and one over which students have considerable control, they are truly critical thinkers.* (Penick, 1989)

⊂ Related *Toolkit98* Activities and Chapters

Not all performance criteria are created equally—they may not all be able to have the instructional effect described in this section. There may be certain design features that make them more or less useful as instructional tools. These features are outlined in **Chapter 2** and **Activities 3.3**—*Performance Criteria—Keys to Success* and **1.9**—*Going to School.*

🕮 Reference Box

J. Penick, J., 1989, "Issues-Based Education—A Need and Rationale," in *Education and Urban Society, 22*(1), pp. 3-8.

Robert Tierney, M. Carter, & L. Desai, 1991, *Portfolio Assessment in the Reading-Writing Classroom,* Christopher Gordon, Publishers, Inc., (617) 762-5577.

BEST COPY AVAILABLE

Relationship Between Classroom Performance Assessment and Large-Scale Assessment

So far, it may seem that we are only discussing classroom performance assessment. But, in reality, the whole enterprise is much more powerful if large-scale and classroom assessment are considered together as a unit. Consider the following points:

1. What if performance criteria for large-scale assessment (which have to be developed anyway) were to be developed by teachers with classroom uses in mind? In order to develop performance criteria, teachers have to engage in the discussions and do the research necessary to develop a common vision of what it means for students to be successful. The effort that goes into large-scale assessment development, therefore, also serves as staff development and contributes to development of instructional materials.
2. What if the performance criteria were uniform across teachers, so that regardless of which teacher observes a student performance, the judgment would be the same? Then assessment credibility improves—and students get the same message across all teachers and grades. Large-scale assessment development can take performance assessment out of the realm of individual interpretation and can promote essential dialogue about what, in fact, constitutes our vision of success for students—our standards.
3. The way we do large-scale assessment sends messages to teachers about what to teach and how to teach it. If large-scale assessment is designed with the instructional end user in mind, there is much more opportunity to influence instruction in desirable and controllable ways.

⊂ Related *Toolkit98* Activities and Chapters

Appendix A—*Sampler* contains many other examples of performance criteria and rubrics.

Activities 2.1—*Sorting Student Work*, 2.2—*How Can We Know They're Learning?* involve sorting student work, generating lists of desirable features, and beginning the process of developing performance criteria.

Activity 2.5—*How Knowledge of Performance Criteria Affect Performance* illustrates the power of rubrics in assisting students to understand the requirements of a task.

298

Activity 2.1
Sorting Student Work

Purposes:

1. To become masters of the targets of instruction (Key 2: Clear Targets) through examining student work for characteristics of quality
2. To experience where criteria come from by "reconstructing" a scoring guide; to see that performance assessment is "do-able"
3. To experience the usefulness, both as staff development and as an instructional tool for students, of the process of developing generalized scoring criteria
4. To illustrate the integration of assessment and instruction—developing performance criteria also increases mastery of targets
5. To motivate teachers to want to pursue activities that increase their mastery of learning targets by showing that it will make instruction faster, easier and better

Uses:

This is an intermediate level activity that can be used in **Chapter 2** to illustrate how development of criteria can help teachers clarify the targets they have for students. The activity can also be used with students for the same purpose. The activity can also be used in **Chapter 3** to demonstrate how to develop generalized criteria, to demonstrate characteristics of quality criteria, and to demonstrate "consequential validity." Prerequisites might include a general stage setting activity (like **Activity 1.1**—*Changing Assessment Practices,* **1.4**—*Seeing Wholes*, or **1.12**—*Assessment Principles*), a defintional activity (like **Activity 1.3**—*Post-it™ Notes*), and a general introduction to the need for performance criteria (such as **Activity 1.5**—*Clapping Hands*, or **1.4**—*Assessment Standards*).

Rationale:

Although overtly about developing scoring criteria, this activity is really about becoming a master of the targets of instruction. This activity develops participants' ability to define essential learning targets for students and identify work that exemplifies different levels of skill on the targets. Teachers nearly unanimously agree that such activities are extremely helpful to them to become better teachers.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead A2.1,O1**—*Sorting Student Work Purposes*
- Set of student work. Several sets are provided in **Appendix B**—*Student Work Samples*. These cover various skills, content areas, and grade levels. Pick the set of most interest to participants. Copy one set (single-sided) for each participant.
- **Handouts A2.1,H1**—*Log Sheet*; **A2.1,H2**—*When Teachers Look at Student Work*[1], **A2.1,H3**—*The Horse Before the Cart: Assessing for Understanding*[2]; **A2.1,H4**—*Build Your Own Performance Criteria* (one copy for each participant). Also pick performance criteria that correspond to the work samples being sorted. For example, if you are sorting math work, you might use the criteria from Kansas (**Sample A.43**) or Illinois (**Sample A.8**) in **Appendix A**—*Sampler*. (Other ideas are provided in the **Activity 2.1 Index** in **Appendix A**—*Sampler*.)
- Plenty of copies of the local content standards document for the content area being sorted

Time Required:

75-120 minutes

Facilitator's Notes:

1. (5 minutes) Explain the purposes of the activity using **Overhead A2.1,O1**—*Sorting Student Work Purposes* and the rationale presented above.
2. (10 minutes) Have participants work in groups of two or three. Pass out the set of student work. First, ask each group to do the performance task themselves—solve the problem, read the passage and answer the questions, sort the fish, etc. (depending on the work samples chosen from **Appendix B**.) If participants are sorting writing, don't ask them to write essays; it takes too long. If the task has a right answer (e.g., math), come to a group (all tables) consensus on what it is. Discuss questions or concerns about the task.

---

[1] Educational Leadership, February 1993, pp. 71-72. Reprinted with permission of the author.

[2] Educational Leadership, February 1993, pp. 22-23. Reprinted with permission of the author.

3. (15-30 minutes) Participants should put the student work into three piles based on its "quality." Describe this task in ways that cue what to do. For example, you can frame the task in terms of grading if it will help participants understand what to do: *Put papers into stacks by what grade you would give to them.* Or, if the use of grading wouldn't appeal to the group, the three stacks could represent levels of student sophistication.

   As participants sort the work, have them *write down the comments they say to themselves* as they place a work sample into one stack or another. (Use **Handout A2.1,H2**—*Log Sheet.*) Some examples of typical comments for math are:

   - **High:** I can tell exactly what the student did at every step. Concise and to the point. Correct answer. Logical and sequential. Good explanation. Precise. Correct labels. Explains thought process. Supports reasoning. Correct answer. Appropriate use of pictures, diagrams, and symbols. Used the right numbers in the right order. Sequencing was purposeful. Evidence that the answer was checked for reasonableness. I could tell exactly what the student did and why he or she did it.

   - **Medium:** Right idea, but computational errors. Forgot a step, but the rest is OK. Didn't go far enough. Didn't clarify. One has to make some inferences as to what the student did. Doesn't show all the steps. Used correct data, but incorrect process. Used correct process, but incorrect data.

   - **Low:** I couldn't follow the sequence. The answer doesn't look reasonable. Used wrong numbers. Used the wrong process. Mixed processes. Shooting in the dark. Illogical. Unclear. No explanation of answer. I had to make a lot of inferences as to what the student did. I was confused by the explanation.

   (Note 1: This level of detail is very helpful both to remind teachers what they meant, and to develop student-friendly versions of the criteria.

   Note 2: Sometimes this procedure leads to very negative phrases at the low end of the scale, such as "clueless," and "totally wrong." One way to avoid this is to suggest that participants think about the comments they would actually make to students when they handed back work—make "I" statements. "I was confused as to what you did," and "I would like to know more about why you used the numbers you did."

   Note 3: Asking participants to place work in stacks to represent "grades" has an added benefit—it clarifies the connection between grades and the features of student work that earns the grade, thereby removing some of the subjectivity from grading.)

4. Circulate and remind participants that the goal is not necessarily to place each paper into the correct pile; the goal is to come up with a good list of reasons for why each paper was sorted as it was. It's the specific comments such as "the student started out well, but seemed to get lost in the middle," rather than the general comments such as, "logical" that will be most helpful. Continually ask participants to "tell me more about that" or "what specific things done by the student were cues that the student used a logical process?"

Continually challenge teachers to think beyond general words such as "voice," "perceptive," and "good communication." Students don't understand what is meant by those words. Since the criteria are supposed to increase student understanding of the features of quality, we need to be clear on the indicators in the work that cue our judgments of quality. Also, challenging teachers to "tell me what indicators in the student work made you judge it logical" causes them to refine their own understanding of what they mean by these terms.

5. (20-30 minutes) When participants are pretty much done, summarize comments on an overhead made from **Handout A2.1,H2**—*Log Sheet*. Write down comments exactly as stated, unless there seems to be a misleading choice of words. In this case write down better wording and ask, "Does this capture it for you?" As groups report out, again challenge participants to think past general terms like "logical" or "understands the problem" to the specific indicators in the work that cued the decision that the response was "logical" or that the student "understood the problem." Emphasize the importance of students knowing what teachers mean by the terms in the criteria.

6. The sorting process used thus far in this exercise is "holistic." Participants end up with a list of comments for high, medium and low performance; any single paper gets only one overall score. Usually during the listing of comments someone will say something to the effect that, "I had trouble placing this paper into one stack or another because it was strong on reasoning but weak on communication." This provides an excellent teachable moment to bring up the **reasons for analytical trait scoring systems**—scoring each student work sample along more than one dimension. The reasons are: (a) it better shows a profile of student strengths and weaknesses, (b) it assists students to be better able to discuss their own work, (c) novices often can't improve all parts of the performance at one time—this helps break the performance down into teachable segments.

   To develop "traits" (important dimensions of performance), you can ask participants to identify comments on the list that seem to go together, and indicate which comments relate to each trait at the high, middle and low levels. For example, people frequently see the traits of communication, problem solving/reasoning, mathematical understanding, and correct computation when sorting math work. Number these traits 1, 2, 3, 4. Then go through the comments they generated and ask them into which category each comment fits. Put the trait number next to each comment.

   For example, the trait of communication (#1) would include, at the high level: I can tell exactly what the student did at every step; good explanation; clearly explains thought process; I had to make very few inferences about what the student did and why he or she did it, etc. I usually color code the trait numbers so that participants can easily see which comments go under each trait.

7. (15-30 minutes) The "so what" discussion. Use this sequence:

   a. Ask participants to compare their list of valued features to local or state content standards that correspond to the type of work sorted. Ask them, "What similarities do you find?" (They always find lots.) This assists participants to see (1) how the content standards are represented in the daily work of students, and (2) that the content standards represent what they already value as outcomes for students; it's not new.

b. Ask participants to look at similarities between their list of valued qualities and the statements in a set of criteria you previously selected from **Appendix A**—*Sampler* that correspond to the content being sorted. There are always many similarities, because, after all, how different can lists of quality criteria be? This enables participants to see that if done well, scoring criteria were not pulled out of a hat; they represent what we all already value in student work; someone has just taken the time to write it down. Point out the alignment between (1) what they value in student work, (2) the content standards, and (3) the way the content standards are being assessed.

c. Ask participants whether they think this would be a useful activity for them to do with their colleagues. (They always answer, "Yes!" because: "It helps to be consistent in grading. It makes me more confident in teaching these skills. It helps us rethink what we really value. It makes us more consistent in instruction. It helps us understand what good work looks like. It helps me focus instruction on what is really important.") The paper **Handout A2.1,H2**—*When Teachers Look...* describes the value to teachers of the process we just completed. (In this paper, the focus was sorting portfolios.)

d. Ask participants whether they think this would be a useful activity for them to do with their students. (They always answer, "Yes!" for the same reasons as above.) The paper **Handout A2.1,H3**—*The Horse Before the Cart: Assessing For Understanding* describes the instructional value of having students help develop criteria for quality.

e. Discuss how the same process can be used for other skills listed in local or state standards such as critical thinking, writing, and reading fluency. Sample scripts:

- **Reading fluency**. Ask teachers to tape students reading over the course of the year. At the end of the year, have teachers place these tapes into 4-5 levels of fluency. Ask them to write down what they are saying to themselves as they place the tapes into piles. They can compare their comments to content standards in reading and a sample reading development continuum—for suggestions see the **Activity 2.1 Index** in **Appendix A**—*Sampler*.

- **Critical thinking**. Ask teachers to choose samples of student work that illustrate different levels of ability to think critically. For each content area, sort the samples into three stacks and write down reasons. Then compare the lists from the various disciplines to determine whether the same criteria might be used across disciplines. You can also try to find where in the content standards critical thinking comes in and compare their lists to sample criteria. Sample criteria are suggested in the **Activity 2.1 Index** in **Appendix A**—*Sampler*.

f. Refer participants to **Handout A2.1,H4**—*Build Your Own Performance Criteria* to review the steps they went through when developing their scoring criteria.

Sorting Student Work

Purposes:

1. To increase knowledge of local or national student learning targets by articulating characteristics of student performance that indicate quality
2. To see where criteria come from by "reconstructing" a scoring guide; to see that performance assessment is "do-able
3. To experience the usefulness of building generalized scoring criteria
4. To illustrate the integration of assessment and instruction

Overhead A2.1,O1

304

Sorting Log Sheet

Category _____	Category _____	Category _____
Sample Numbers:	Sample Numbers:	Sample Numbers:
List reasons why:	List reasons why:	List reasons why:

Handout A2.1,H1

305

306

When Teachers Look at Student Work

Educational Leadership 50, 5:71-72. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1993. All rights reserved.

Christine Sobray Evans

In California schools, more and more teachers are relying on portfolios to put authentic student work at the center of education.

When I started teaching 17 years ago, I would look at student work only to give it a grade and check it off. I relied on publishers to test my students and to tell me what to teach. I would start at Chapter 1 and hope to finish the text before the end of June. In those days, my focus was on me—my plans, my lessons, my teaching. It embarrasses me now to confess this. I would like to call all of my early students and apologize for my selfishness.

Since that time, a revolution has taken place in my classroom and in classrooms like mine across the state. In California, we have moved beyond a skill-based curriculum, which has to do with teaching, to a meaning-centered curriculum which has to do with learning. Because of this shift, student work has become the center of education. Teachers no longer follow the guidelines of publishers but, instead, look at their students to determine what must happen next.

The Power of Portfolios

A portfolio is one strategy teachers can use to focus on student work. In Poway, we've been using Language Arts portfolios for almost three years now. That might seem like a long time in education, but we are just beginning to reach new understandings.

At first, we coped with the physical problems of portfolios. Who should keep them? Where to keep them? How often should we look at them? What should go inside? Since our new report card stated that "grades" were based on a collection of work, portfolios appeared in some form in almost every elementary classroom.

That first year, portfolios were much like the work folders that teachers had traditionally kept for students. Teachers, rather than students, made most of the selections. The portfolios were organized storage bins, where work went in but never came out. Then at the end of the year, the portfolios were sent home—much like a scrapbook for the year.

In a few places, however, some exciting things were done with portfolios. At Los Penasquitos Elementary School, the 3rd grade team met in a tiny storage room—portfolios in hand—to look more closely at the 3rd grade language arts program. All the portfolios were shuffled and dealt out like a deck of cards. We then spent about 45 minutes, silently reading and making notes. The first surprise was how accurately the portfolios portrayed our students. I remember reading through one folder saying, "Mary, Sarah appears to be struggling with writing. All of her pieces are short and done just to complete the assignment. Her reading list shows that she enjoys reading grade-level books."

"That's right," Mary replied, surprised that I knew so much about a student I had never met. This interchange was repeated time and time again. It gave us great confidence that portfolios could provide an undistorted image of student effort and achievement.

When we look at the collection of portfolios as a whole, we made some important discoveries, too. Worksheets and formula writings were much less informative than student-generated topics. Our 3rd grade team decided to spend less time on formula writing assignments that asked students to fill in the gaps—and more time teaching young writers to focus on one topic. So, although we were really looking at *student* learning, we learned a tremendous amount about our students and our 3rd grade program.

Developing Models and Rubrics

The power of looking directly at student work as a *team* cannot be overstated. Real student work gives teachers a starting point fo

Handout A2.1,H2, p. 1

BEST COPY AVAILABLE

conversations that get to the essence of what happens in classrooms. Samples of student work are concrete demonstrations of what is known and what is not known. They also provide teachers with signposts that mark how far we've come and point us in the direction we must follow.

At Los Penasquitos, the principal provided released time for the grade level teams to get together. However, in schools where there is staff interest, teachers can still meet. Bernardo Heights Middle School is a perfect example. Last fall teachers on the 7th and 8th grade teams decided to establish schoolwide standards in writing. Using a prompt from the California Assessment Program (CAP) handbook, we asked every 7th and 8th grader to write about a school-related experience. Students completed these essays during a 45-minute class period—all were first draft pieces.

That afternoon we met to select examples of student writing to serve as models for our 6-point rubric. Papers were shuffled about from stack to stack. We placed pieces that showed promise in the *high* pile, papers that looked like typical responses in the *middle* pile, and papers that raised concerns into the *lower* pile.

In the high pile, we hoped to find "outstanding" examples that would rate a 6 and "good" ones that would rate a 5. On the chalkboard, we listed characteristics that put a paper into the top category: selecting a dramatic topic, using precise vocabulary, and creating sentence variety, to name a few.

"I think this might be a 5," said Kathy. She read it to our tiny group and then argued for it: "It has more sophisticated sentence structure than we normally see. The vocabulary is strong, and the details paint a clear picture. I know it's not a 'wow,' but it's clearly a 5."

In California, we have moved beyond a skill-based curriculum, which has to do with teaching, to a meaning-centered curriculum, which has to do with learning.

Our dialogue continued until we had models for each point and had annotated each piece with information for future references. Each teacher in the group received copies of the models and annotations, and we've shared them with students and parents. We wanted everyone to have a clear understanding of how we look at student writing.

Spotlight on Students' Work

To our surprise, the models were just one benefit of this entire process. More important, we felt, were the discussion and negotiation that occurred that afternoon. In addition to learning about student writing, we valued the opportunity to talk to one another about the work we do. That sense of being a community of learners continued for the entire year.

Our "community of learners" has since expanded to students and their parents. We use our models and rubrics to train young writers to examine their own work and that of their peers:

- Can you find examples of precise vocabulary?
- Can you play a videotape in your head and see the action?
- Here is a paper written last year.

You said it was a 3. Can you work with your group to make it a 4?

Questions like these help us to zero in on the elements that make writing come alive for the reader.

Parents have shown their support by attending evening workshops so that they, too, can learn to look at student work through a new lens. "This is not easy," one mom wrote after a training session on rubrics. "However, I think that I can help my daughter better now."

Teachers, students, parents—we are all focusing on real work to give us insights about learning. Sometimes, though, these insights make us uncomfortable. "I'm going to have to re-do mine," Tom confided to me, "I just didn't know what *good* looked like!"

The portfolios that fill our crates and cabinets, and sometimes cover my dining room table, are no longer scrapbooks of completed assignments. Instead they are showcases for the best work a student can do. They allow young people, like Tom, a chance to reflect, revise, and re-start, if necessary, in order to paint a portrait of themselves as learners. Portfolios also give me, the teacher, the same opportunity to reflect, revise, and re-start. I no longer look to publishers to tell me what to do next. My students show me the way.■

Christine Sobray Evans is a teacher at Bernardo Heights Middle School, 12990 Paseo Lucido, San Diego, CA 92128.

Handout A2.1,H2, p. 2

308

The Horse Before the Cart: Assessing for Understanding

Educational Leadership 51, 5: 22-23. Reprinted with permission of the Association for Supervision and Curriculum Development. Copyright 1994. All rights reserved.

Rebecca Simmons

If students are to develop understanding, assessment must be an ongoing part of instruction.

Understanding is a tricky thing to get one's mind around. We want students to be able to employ knowledge in flexible and novel ways, to develop coherent networks of concepts, to use what they learn in school to understand the world around them, and to develop an interest in lifelong intellectual pursuits. But to help students achieve such understanding is no mean feat.

To teach for understanding requires that we attend closely to assessment. We cannot assume that because we have taught a lesson "well," students have understood it. Rather, we have to seek evidence of understanding through student performance. Going to the heart of the matter, one science teacher said that he could do amazing things in the science classroom all year long and, at the end of the year, his students would have seen amazing things. Beyond that, without clear-cut student demonstrations of understanding, he could not be sure what students had gained from his class.

Assessment is not something that we tack onto learning; it is an essential *ongoing* component of instruction that guides the process of learning. Ongoing assessment uses exhibitions, student explanations of concepts, the writing of a poem or a song, or any number of other thought-demanding performances to evaluate and reflect on students' work. From the beginning of our curriculum planning, we need to identify these understanding performances—the ways in which we expect students to demonstrate their understanding. In a sense, ongoing assessment is the horse that leads the cart of understanding.

Using Assessment to Build Understanding

One teacher who collaborated on the Teaching for Understanding Project at the Harvard Graduate School of Education designed several understanding performances that provided the basis for ongoing assessment for a high school literature unit. Based on the generative topic, "Trials and Literature: Who Determines Justice and How?" the unit included works of literature such as *Mutiny on the Bounty*, *To Kill a Mockingbird*, *Inherit the Wind*, and *Twelve Angry Men*.

As students studied each piece of literature, they kept charts in their notebooks to identify where justice-related themes were introduced and how those themes were developed. Students shared these charts to compare viewpoints and evidence for arguments. The charts also formed the foundation for more formal debates about the outcomes of particular trials.

In addition to the charts, students kept journals in which they recorded their responses to and questions about their reading. Later, in large-group discussions, the teacher encouraged students to wrestle not only with the answers to their questions but to think through *how* they were answering them. Which questions could be answered definitively? Which might have several answers? Were some answers better than others? As a culminating work, students wrote papers in which they chose one aspect of justice and used two or three of the texts they had to read to defend their thesis.

This example illustrates several important points concerning the methods and utility of ongoing assessment. First, the teacher helped students to understand from the outset the standards by which their work would be judged. For example, throughout the discussions, journal writing, debates, and papers, students knew that they needed to use evidence from the text to support arguments, rather than arguing from personal opinion. Where possible, they were to use metaphorical evidence as well as literal evidence from the text, a challenging but critical task in developing a deep understanding of literature.

Second, students kept track of their work throughout the course of the unit. This "record of understanding" served a variety of assessment purposes. It provided concrete evidence of work, evidence that formed the basis for ongoing feedback to students about their understanding. Students used their work-in-progress as a springboard for self-assessment and reflection on their products and processes. They also used their record to give and receive peer critique. Informal peer critique occurred during discussions arising from the student journals and as students shared their notebook charts. More formal peer critique took place in the small group debates.

Handout A2.1,H3, p. 1

In addition, the record of understanding enabled the teacher to adjust the curriculum to better meet student needs as well as provide students with ongoing individual feedback and evaluation.

Third, students came to understand through the constant process of performance and feedback that some questions are not easily answered, but must be grappled with as best one can.

Finally, the culminating papers on justice served as an elaboration of the understanding performances the students had been building, demonstrating, and consolidating throughout the unit.

No More Surprises

Making the standards for good work clear to all is implicit in a classroom where students and teachers take time to reflect on and assess understanding performances. Students should know from the start the standards they are working toward. Equally important, various benchmarks of good performance can help students gauge their own understanding. Students should *not* be surprised by the grades they earn.

In addition, students cannot achieve deep understanding if they receive evaluation passively. Taking time and energy to reflect on and improve one's work are essential to the understanding process itself.

Both teachers and students are often wary of making criteria for good work public and involving students in assessing their own and their peers' work. In our research we have found that some teachers and students fear that students do not have the ability to assess themselves. Some students are reluctant to engage in peer- and self-assessment because they feel that such activities are the teacher's—the expert's—job and that their own work will ultimately suffer from their own or their peers' assessment. Yet, we have seen in our work that when students and teachers set forth criteria and use them as a basis for reflection on student work, both groups are often surprised at the positive outcomes.

One of our teacher-collaborators experimented with using public criteria in a unit he developed on global warming. He made the criteria for evaluation an integral part of the instruction, giving students guide sheets and discussing them with the students. These criteria included accuracy in carrying out and reporting one's work, clarity in scientific discussion and writing, appropriate use of scientific vocabulary, appropriate scientific methodology, and sound reasoning based on data and coherent arguments.

At the end of the unit, the teacher was pleased and somewhat surprised at the students' attention to the criteria. Students kept the guide sheets with them as they worked through all activities, often referring to the criteria to see if their work was meeting the proper standards. The criteria were seen as a useful tool for thinking through the quality of their work, rather than as a burden.

The public airing of criteria also empowers students. As students begin to see that they do have the capability for peer- and self-assessment, they feel a new sense of control over their work. As teachers begin to see that students can engage in peer- and self-assessment, they, too, become excited by what students can accomplish. They may also feel relieved of the burden of being responsible for every aspect of ongoing assessment in the classroom.

Back on Track

Ongoing assessment is also useful in getting students back on track when they have problems with a performance. In a geometry unit, students were to design the floor plan of a hypothetical community center as a means of exploring the area of geometrical constructions. Early in the project, student groups presented their progress. Much to the surprise (and chagrin) of the teacher and researcher who had co-constructed the unit, some groups were using the formula for perimeter rather than area in their calculations. They had gone back to the familiar—perimeter—and had forged ahead enthusiastically in the wrong direction.

One might be tempted to dismiss this as a trivial error, but this sort of error happens all the time, especially in mathematics and science where rote memorization of formulas is often the norm. Uncovering this glitch through ongoing assessment provided the teacher with an important teaching opportunity, the chance to make sure that students in the class were not just using formulas by rote, but were thinking through what they needed to do and why. By discussing this error early on, the students were able to progress in their work and ultimately succeed in their projects.

This back-and-forth between performance and ongoing assessment illustrates the essence of understanding; it is not the smooth all-or-nothing ride that students often consider it to be. Rather, developing understanding involves a series of bumps and starts that emphasizes the importance of the processes and the developing products of understanding rather than the importance of any one grade.■

Rebecca Simmons is a Research Associate for Harvard Project Zero and the Project Manger for the Teaching for Understanding Project, Harvard Graduate School of Education, 323 Longfellow Hall, Appian Way, Cambridge, MA 02138

Handout A2.1,H3, p. 2

BEST COPY AVAILABLE

Build Your Own Performance Criteria
Sample Topic: Self-Reflection Letters

Here's how you build generalized performance criteria. I will illustrate the steps using student self-reflection as an example. Remember, we are *not* developing these performance criteria so we can "grade" student self-reflections. We are doing it as an exercise in clear thinking about what good self-reflection looks like.

Step 1: Gather samples of student performance

Gather samples of student performance that you feel illustrate the skill or behavior in question. In our case we are trying to assess student self-reflection skills, so we might collect things that students have written in answer to questions like: "Select a piece of work that you feel illustrates your best effort. Why did you select this piece?" "What does the piece you selected show about you as a __________?" (The blank would be filled by: writer, math problem solver, life-long learner, or whatever skill students are self-reflecting about.) Other questions that might elicit student self-reflection might be: "How have you changed this year in your ability to _______?" or "What do you currently understand about _________? What do you still need to understand?"

Step 2: Sort student work into stacks; write down reasons

Place the samples of student work into three piles: strong, middle and weak. For our example we might have three piles that illustrate levels of sophistication with respect to self-reflection. As you sort the student work, write down your reasons for placing pieces in the various stacks. In other words, if you place a piece in the "sophisticated" pile, why? What cues you that the work is sophisticated? What are you saying to yourself as you place a piece of work into a pile? What might you say to a student as you return this work? These reveal criteria.

Keep sorting work until you are not adding anything new to your list of attributes. Try to create as large and diverse a list as possible. Recent lists have included such things as:

High: detailed, many things covered, insightful, self-revelation, examples provided, the student seemed motivated, sets goals for the future, looks at more than one thing, considers content as well as process, takes risks, accurate, discussion is related to criteria, growth supported with examples, organized well, sincere, honest, there are comparisons over time, shares feelings, looks at both strengths and weaknesses, there is depth to the analysis, there are good reasons and explanations, revealing, voice, easy to read, looks at skill improvement, there is ownership, there is a personal reaction, it is specific, it looks ahead, it is thorough, ideas are synthesized, it is readable, it is neat, ...

Handout A2.1,H4, p. 1

Middle: shows then hides, beginning of ownership, two-dimensional, some specifics, describes performance but leaves a lot out, few insights, focuses on only a few things, considers content or process but not both, is somewhat accurate, doesn't seem completely honest, there are descriptions of individual pieces of work but no comparisons over time, I have to make some inferences as to what the student meant, ...

Low: Vague, simple restatements, mechanical, focuses mainly on surface features, obvious, same old same old, no examples, purposeless, I like it/I don't like it, one-dimensional, superficial, does not seem to be aware of the need to set goals, takes no risks, doesn't seem honest, is not accurate, there seems to be no ownership, I can't follow what the student is trying to say, hard to read, not organized well

Step 3: Cluster the reasons into "traits" or important dimensions of performance

Cluster similar attributes together and begin to form important "traits" or "dimensions" of self-reflection. For example, the above attributes *might* be clustered into the traits of: **SKILL ANALYSIS, SINCERITY, GOAL SETTING,** and **PRESENTATION.** Number each trait and place next to each comment in the high, middle, and low categories the number of the trait to which it relates.

Make sure that your "traits" cover everything of importance. For example, **SKILL ANALYSIS** at the HIGH level might include: detailed, many things covered, insightful, examples provided, looks at more than one thing, considers content as well as process, accurate, discussion related to criteria, growth supported with examples, looks at both strengths and weaknesses, depth of analysis, revealing, specific, thorough, and synthesized ideas.

Be prepared for frequent changes, at least at the initial stages of building the criteria. Ideas have to have time to gel. We get new perspectives and insights as we attempt to "make sense" of our criteria.

Step 4: Write a value-neutral definition of each trait

Write a definition of each trait. For example:

> *Sincerity reflects how serious the student is about the self-reflection process. It is the degree of ownership, effort and honesty in the self-reflection. It is the degree to which the student makes a real effort to self-analyze.*

Handout A2.1,H4, p. 2

These definitions should be "value neutral"—they describe what the trait is about, not what good performance looks like. (Descriptions of good performance on the trait are left to the "high" rating.) Here is an example of the definition above converted to a statement that is *not* value-neutral: *Sincerity is taking ownership for work, really trying hard and being very honest; the student has made a sincere effort to self-reflect.*

Step 5: Find samples of student performance that illustrate each score point on each trait

Find samples of student work which are good examples of strong, weak and mid range performance on each trait. Be sure to have several sets representing typical ranges of performance at each level of: early elementary, late elementary, middle school, and high school. These can be used to illustrate to students what to do and what "good" looks like.

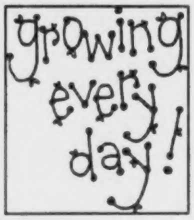
Step 6: Make it better

Criteria evolve with use. Try them out. You'll probably find some parts of the criteria that work fine and some that don't. Add and modify descriptions so that they communicate better. Choose better sample papers that illustrate what you mean. Revise traits if you need to. Let students help—this is a tool for learning.

Handout A2.1,H4, p. 3

Activity 2.2

How Can We Know They're Learning?



Purposes:

1. To learn a technique (examining samples of student work) for examining valued learning targets for students
2. To broaden understanding of the importance of clear criteria for quality assessment
3. To contribute to the criteria that can be used to assess students' work
4. To broaden understanding of assessment techniques

Uses:

This introductory activity for parents and community members can be used in **Chapter 1** as an introduction to changes in assessment and instruction or to illustrate valued learning targets for students. In **Chapter 2** it can be used to show how assessment can integrate with instruction. In **Chapter 3** it can be used to demonstrate how criteria are developed.

It has been used as part of a series of sessions with parents over time that builds a common understanding using "neutral" student work, and then gradually moves to examination of anonymous samples from the same school, and finally to examination of the work of their own children once the language of assessment has been established, and agreement on qualities to look for in the work are clear. The sessions often include teachers and students.

Rationale:

This activity, an adaptation of **Activity 2.1**—*Sorting Student Work*, is intended for parents and is centered on the question "How can we know they're learning?" As schools and classrooms expand the use of alternative and performance assessments, it is vitally important that parents are invited to become an integral part of the conversations about student learning. In the past, such conversations have focused on brief, often numeric, summary information that does not provide parents with insights into the richness of learning—the expanding knowledge, skills, and capabilities of children.

Using the student math work in **Appendix B (Sample B.4**—*Camping Trip*), this activity provides an opportunity for parents and family members to expand their knowledge of alternative assessment and to become involved more fully in rich and valuable conversations about children's learning. In addition, this activity includes responses from a mathematics educator to enable parents to look inside the reasoning that teachers use when assessing student work.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens, or chart paper and pens
- Eight to 10 samples of student math work representing different levels of sophistication—the set is provided in **Appendix B**—*Student Work Samples* (**Sample B.4**—*Camping Trip*)
- **Overheads A2.2,O1**—*How Can We Know They're Learning Purposes*; **A2.2,O2**—*Assessment Is About...*; and **A2.2,O3**—*Holistic and Analytical Trait Scoring*
- **Handouts A2.2,H1**—*The Redwoods*, **A2.2,H2**—*Fox*, and **A2.2,H3**—*Waialae's Draft Standards and Indicators for Creative Problem-Solving* (one set per group of two-three participants); other generalized math problem-solving scoring guides could also be used (see suggestions in the **Activity 2.2** Index in **Appendix A**—*Sampler)*

Time Required:

Part A: 1 ¼ to 1 ½ hours; Part B: 30-40 minutes

Facilitator's Notes:

A. Basic Activity

1. (2 minutes) Introduce the purpose of the activity, as needed, with **Overhead A2.2,O1**—*How Can We Know They're Learning Purposes*.
2. (3 minutes) Open the session with **Overhead A2.2,O2**—*Assessment Is About...* and use it to set the context for the activity: *There are other, more important purposes for assessing student achievement than merely to grade. Assessment of student achievement is most productive when it is used as a tool to communicate about student achievement and to plan further instruction.*

3. (20 minutes) Provide a mini-practice that emphasizes how to describe the qualities of student work. This mini-practice uses samples of student writing because most parents feel comfortable with their ability to comment about the quality of writing.

   a. Begin with the piece that represents a relatively strong piece of student writing (**Handout A2.2,H2**—*Fox*). Read the paper aloud because writing plays to the ear differently than it might to the eye. Ask the group to review the piece and identify its strengths; key qualities that make it work. Give them time to think about this individually and then in small groups.

      Bring the whole group back together and record participant responses on one side of a blank transparency.

   b. Then read and discuss **Handout A2.2,H1**—*The Redwoods*, a weaker piece of writing, in the same manner as above. In the whole group, record participant comments on key qualities on the other side of the transparency.

   c. Summary: *We've done this activity with thousands of parents, teachers and students and they all come up with pretty much the same list of characteristics that make writing work. We tend to value the quality of the ideas in a piece of writing* (indicate the statements they made that relate to the quality of ideas); *how well the ideas are organized* (again indicate their comments that relate to the trait of organization), *the appropriateness of the style for the audience and purpose* (indicate comments that relate to voice), *the words that are chosen to convey thoughts* (indicate comments that relate to word choice), *the fluency and correctness of the sentences* (indicate comments that relate to sentence fluency), *and the correctness of the grammar, spelling, punctuation, capitalization and paragraphing* (indicate comments that relate to conventions).

   (If you need assistance relating parent comments to the traits listed above, see **Sample A.10**—*Six Trait Model* in **Appendix A**—*Sampler*.)

4. (5 minutes) Now you'll ask participants to do the same thing for math—look at samples of student math solutions and list characteristics of the solutions that make them work. Bridge to math problem solving by noting that teachers, too, are engaging in conversations about the quality of student work in areas besides writing; they are especially looking at the big ideas that are central to learning for all children—like problem solving. (If there is a school vision in place, note the connections between the vision and students' becoming effective problem solvers.) If one purpose of this activity is to connect assessment and national standards, identify the NCTM mathematics problem-solving standard as one which cuts across ages and grades.

   Have participants move to groups of two or three. You might, for example, have parents add the ages of their children/grandchildren and find two other people close to their own total.

5. (10-20 minutes) Pass out the student math solutions (from **Appendix B**, **Sample B.5**). Ask parents to place this work into three piles based on the "quality" of the solution. Describe this task in ways that cue what to do. For example you can frame the task as dividing the work into levels of sophistication, using language such as "beginning work", "developing work that shows both strengths and weaknesses", and "strong work." Or it might be helpful to cast participants as those responsible for grading: "Put papers into stacks by what grade you would give them." (Note that using grading as the prompt often results in questions and discussion of current grading practices; this is great when it's intended, but needs to be anticipated so that the activity doesn't lose its focus.)

   As participants sort the solutions, they should write down the basis on which they placed papers in each stack. *You thought the papers in this stack were better than the papers in this other stack. Why? What features of the work did you like? Not like? What makes the work in this stack different from the work in that stack?* One strategy is to designate a person to be the observer/recorder for each group whose responsibility is to capture the evaluative language used, and the reasons cited for placement. (When there are teachers present, this is often their role so that parents do not defer to their judgment and limit their own discussion.)

   The goal is not to score or "grade" all the papers, but to articulate reasons for placing samples into stacks. Therefore, participants only continue sorting and discussing until all their reasons are listed. Remind each group of this as they work.

6. (10 minutes) Put up a clean overhead transparency (or use chart paper) with three columns labeled "Strong, developing, and beginning" or "high, middle, and low." (Use whatever titles will appeal to the group.) Have groups tell you the placement of each paper. Start a fourth column with a "?" heading. If there is a disagreement about placement, write the paper number under the "?" heading.

   There is no "right" answer for this sorting. The goal is for participants to articulate what they value, come to consensus, and be able to justify their judgments using consistent terminology.

   List all the criteria the parents gave for their sorting.

7. (5-10 minutes) Try to resolve placement of the "?" papers. Have the various groups describe their rationale for placement. Add their reasons to the criteria list. The goal is to have a rich and elaborated list of criteria that cover many features of what is valued.

   Depending on time, you can use disagreements as an opportunity for parents to clarify and articulate what they saw in particular papers. This is also an opportunity to point out that while it is possible and important for educators to come to agreement on key qualities of student problem solving, there are also differences among what individual teachers value. The point can be made that making what is valued overt and clear to students is critical to quality assessment.

8. (5-10 minutes) Using **Handout A2.2,H3**—*Waialae's Draft Standards and Indicators for Creative Problem-Solving*, ask parents to look for similarities between their list and the list in the handout. For most groups, their own careful look at student work resulted in the identification of quite similar key dimensions or traits of quality. (Also look at some actual performance criteria used to score open-ended math work. Suggestions are given in the **Activity 2.2** Index in **Appendix A**—*Sampler*.) Points to make:

   a. Ideas of what to look for in student work to judge its quality are quite similar across the country; teachers, parents, and students tend to agree on the features that make solutions to math problems work. This is true for other student skills as well, such as writing, oral presentations, and how to do science experiments.

   b. The process they have used is one that professional educators also engage in as they seek to clearly define assessment targets and clearly articulate characteristics of quality work.

   c. This careful examination of work is the basis for performance assessment; some of the learning targets we have for students require humans to carefully examine student work and judge its quality. The processes we have just gone through help make this judgment process consistent across teachers.

9. (5-10 minutes) Close with oral or written parent reflections. Some prompts might be:

   - What insights did you gain?
   - What questions remain?

B. Extensions

1. (20 minutes) The session can be extended to provide awareness of the benefits and drawbacks of different kinds of scoring. You might pose the question, *Did you see some quality or qualities in a work that made you want to place it in one stack, and then other qualities that seemed to place it in a different stack? What were they? How do you feel about the forced choice of one stack or the other?* Note that the forced choice that they used echoes holistic scoring in which the overall impression of the work guides scoring.

   Show parents sample holistic and analytical trait performance criteria (rubrics). Suggestions are provided in the **Activity 2.2** Index in **Appendix A**—*Sampler*. First have them verify that the list of qualities assessed in these rubrics are the same as those identified in Part A of this activity.

Use **Overhead 2.2,O3**—*Holistic vs. Analytical Trait Scoring* to clarify the distinction between holistic and analytical trait scoring. Ask the groups to discuss when each kind of scoring would be useful. What are the positives for parents and students if student work is scored analytically? What are the positives for parents and students if student work is scored holistically? What are the negatives of each?

This can also lead into discussion of what kinds of reporting helps parents understand their children's learning and gives them some pointers about what's worth praising in their child's work and what areas need more attention.

2. (10-20 minutes) Another extension is to try scoring student work Take one or two new solutions to the same problem, or a sample of student responses to another problem solving task, and try to score it analytically with a math rubric from **Appendix A**. (Suggestions are provided in the **Sample 2.2** Index of **Appendix A**—*Sampler*.) As before, parents should justify their judgments using phrases or descriptors from the scoring guide.

How Can We Know They're Learning?

Purposes:

1. To use samples of student work to examine valued learning targets for students
2. To broaden understanding of the importance of clear criteria for quality assessment
3. To contribute to criteria used to assess student work
4. To broaden understanding of assessment techniques

Overhead A2.2,O1

Assessment Is About the Exchange of Information

Assessment *is* about:

- Providing direction for next steps in learning
- Exchanging information with students and parents about student achievement & progress
- Describing where students' work is now, on the way toward important learning targets
- Helping students set learning goals and plan next steps toward the targets

Assessment is *not* about:

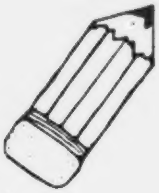
growing every day!

- Saying to students, "Bad luck; you got it wrong."
- Punishing students
- Labeling students and limiting their expectations of themselves

Overhead A2.2,O2

Holistic Scoring

Scoring that responds to a student's work as a whole; assigning a single score or description of quality based on clearly defined criteria.



Analytical Trait Scoring

Scoring that responds separately to each of the key qualities or dimensions of the student's work; assigning individual scores or quality descriptions for each dimension. These dimensions are called traits.

Overhead A2.2,O3

The Redwoods

Last year, we went on a vacation and we had a wonderful time. The weather was sunny and warm and there was lots to do, so we were never bored.

My parents visited friends and took pictures for their friends back home. My brother and I swam and also hiked in the woods. When we got tired of that, we just ate and had a wonderful time.

It was exciting and fun to be together as a family and to do things together. I love my family and this is a time that I will remember for a long time. I hope we will go back again next year for more fun and an even better time than we had this year.

Handout A2.2.H1

FOX

I don't get along with people to good, and sometimes I am alone for a long time. When I am alone, I like to walk to forests and places where only me and the animals are. My best friend is God, but when I don't believe he's around sometime's, my dog stands in. We do every thing together. Hunt, fish, walk, eat and sleep together. My dog's name is Fox, 'cause he looks like an Arctic Fox. Fox and I used to live in this house with a pond behind. That pond was our property. The only thing allowed on it (that we allowed) was ducks & fish. If another person or dog would even look like going near that place, Fox and I would run them off in a frenzy. There was a lot of rocks around, so I would build forts and traps for any body even daring to come near. The pond had a bridge that was shaded by willows, so on a hot day me and Fox would sit on that bridge & soak our feet, well, I would soak my feet, Fox just kinda jumped in.

At night, the pond was alive with frogs, so I would invite this kid over, (he was a guy like me) and catch frogs. After we had a couple each, we would pick the best looking one out of our group and race them. The winner gets the other guys frog.

In the winter, the pond would freeze over, and I got my iceskates out. The pond was now an ice skating rink. Fox would chase me as I went round & round the pond.

After about a year, I was riding my bike patroling the area around the pond. With Fox at my side, I raced downhill toward the pond. I tried to stop, but my back tire went into a skid. I went face first into murky, shadowy waters. When I went down, a minute later I felt something pull on my shirt, I grabbed it, not knowing what to think, when I hit the surface, I saw that it was Fox, pulling on my shirt as if he was trying to save me. He was to little to save me if I was really drowning, but it was the thought that counts, I owe him one.

Another year passed. One day my mom got home from the store, and she bought me a rubber raft. It was just a cheap one, but it was mine. I blew it up with a tire pump. It was just the right size for me & Fox. Out of respect for Fox, I named it the USS Fox and christened it right in the pond.

On sunny days, I would take the raft out & lay in the sun with Fox on my legs. One day, when I was asleep in the raft, the wind blew pretty hard and blew my raft right into a bunch of sticks and rocks, the USS Fox was given a sad salute, and then was no more.

Another year passed, and this would be our last year by the pond. I admired and respected that pond more than I ever did that year. But, at long last, all good things must come to an end, we moved to another town. Fox & I still visit the pond, but it'll never be like them 3 years when she was mine.

Handout A2.2,H2

Waialae School
Draft Standards and Indicators for Creative Problem Solving
Draft #3

Definition of Creative Problem Solving

Creative problem solving includes the ability to understand the problem, build solutions and follow through, and produce an outcome which addresses or solves the problem. It includes an element of creativity which requires that children extend beyond their own limit in the problem solving process or outcome and involves the ability to see a multitude of possibilities. Creative problem solving is developed through continuous opportunities to perform authentic tasks and involves children being able to construct meaning or find connections within the larger context of real world situations/problems.

Standard #1: Understands Problem

- Distinguishes between relevant and irrelevant information and factors in the problem
- Links experience and prior knowledge to problem
- States the problem in a question or series of questions
- Reflects on possible multiple approaches to arrive at outcomes and/or possible multiple outcomes for the problem
- Relates problem to real world situations and contexts
- Makes meaning of the problem in personal, social or community sense
- Interprets problems from other points of view and multiple perspectives

Standard #2: Building Solutions and Following Through

- Making a Plan
    - Draws on and applies past knowledge
    - Identifies information and resources
    - Evaluates multiple approaches
    - Selects appropriate, feasible, efficient, simplest solution strategy
    - Describes plan of action
- Following Through
    - Accesses and gathers appropriate materials and data
    - Takes risks by trying ideas that are new or original for the child
    - Evaluates appropriateness of strategy
    - Adjusts strategy based on evaluation
    - Explains and describes what worked and did not work during process to create final product

Standard #3: Outcomes

- Originality
  - Product is a stretch of child's limits
  - Product is unique for child
  - Uses alternative resources/materials
- Quality
  - Product is clear and coherent
  - Product is accurate
  - Product is refined and polished
  - Product is organized
  - Demonstrates effort
  - Demonstrates persistence
- Appropriateness
  - Product/process is related/connected to problem or task
- Reflection
  - Evaluates product and makes suggestions for improvement

Standard #4: Creativity

- Flexibility in Thinking
  - Generates multiple approaches
- Transfer in Thinking
  - Applies problem-solving approaches to new situations and contexts
  - Communicates verbally and nonverbally about application of concepts to new or real world situation
- Originality
  - Demonstrates divergent thoughts and approaches which are different from past thoughts and practices for that child
  - Demonstrates divergent thoughts and approaches which are unique for all
  - Reflects on the uniqueness of thoughts and approaches

Handout A2.2,H3, p. 2

Activity 2.3

Ms. Toliver's Mathematics Class

Purposes:

1. To discuss and examine what makes assessment effective in the classroom
2. To illustrate the term "authentic" assessment
3. To think about assessment beliefs and principles
4. To observe a teacher who models good assessment practices

Uses:

This activity is at an intermediate level. It can be used in **Chapters 1, 2,** or **3** to illustrate terminology, illustrate continuous monitoring of student achievement, demonstrate good quality assessment tasks that are integrated with instruction, or stimulate teacher reflections on effective classroom assessment. Prerequisites might be **Activity 1.12**—*Assessment Principles*, and **1.5**—*Clapping Hands* or **1.11**—*Assessment Standards*.

Rationale:

This activity will help participants view assessment more as a process over time than as an activity or test given to students at the end of instruction. Teachers need to see examples of how it looks in the classroom when a teacher is engaged in continuous and standards-based instruction and assessment. On this video, Kay Toliver is featured interacting with students at East Harlem Tech in New York City. After watching the video, discussion groups are formed to look at the "big picture" of effective classroom assessment practices.

This activity can be used to bring many aspects of the classroom together and to emphasize assessment as a vital part of instruction and learning. Good instruction requires knowing:

- The destination (targets) of what students know and are able to do
- How to provide multiple and various opportunities for student practice
- How to create authentic instructional experiences for students
- How to provide good feedback to students so they know what to do next to continue to improve their academic achievement

- One's own beliefs or principles guide assessment practices

Materials:

- Video, *Kay Toliver and the Students of East Harlem Tech*
- **Overheads A2.3,O1**—*Ms. Toliver Purposes*, **A2.3,O2**—*What Makes Assessment Effective in the Classroom?*, **A2.3,O3**—*What Is "Authentic Assessment?"*, **A2.3,O4**—*Authentic Assessments Can Include Many Formats*, **A2.3,O5**—*A Question Might Be*, **A2.3,O6**—*A Closer Look at Assessment Principles*
- **Handouts A2.3,H1**—*What Makes Assessment Effective in the Classroom?*, **A2.3,H2**—*Assessment Principles*

Time Required:

$1\frac{1}{2}$ to $2\frac{1}{2}$ hours

🕮 Reference Box

Good Morning Ms. Toliver is a Peabody Award-winning video, VHS, color, stereo, 27 minutes long. The program is underwritten by the National Science Foundation, ARCO, the U.S. Department of Energy, and the Corporation for Public Broadcasting. It is available for $24.95 plus $6.50 UPS shipping (1997 prices) from FASE Productions, 4801 Wilshire Blvd., Suite 215, Los Angeles, California 90010, (800) 404-3273, fax: (213) 937-7440, e-mail: inquiries@fasenet.org

Facilitator's Notes:

1. (10-15 minutes) Go over the purposes for the activity (**Overhead A2.3,O1**—*Ms Toliver Purposes*) and give a pre-video overview using **Overheads A2.3,O2** to **A2.3,O6**.

   a. Using **Overhead A2.3,O2**—*What Makes Assessment Effective in the Classroom?*, describe effective classroom assessment practices:

      - Knowing the destination (learning targets) for students. This means more than just being able to say, "I want my students to be critical readers, good writers, and know how to work well with others." A teachers *knows* the destination when, for example, he or she can define what critical reading is and identify examples of different levels of sophistication in the day to day performance of students. A teacher is a master of the targets of instruction when his/her students can also define the outcome and identify examples of different levels of sophistication and their own progress toward it.
      - Providing opportunities for practice on these learning targets.

- Having clear purposes for the assessment that go beyond grading. Most classroom assessment should be for the purpose of providing feedback on progress toward our clearly defined targets. The most effective assessment is student self-assessment; but only if students understand the nature of quality and how to systematically apply criteria of quality to their own work.

Ask participants for other practices.

b. Ask participants to compare and contrast "authentic" and "alternative assessment." Participants often get authentic and alternative definitions confused. Alternative assessment is anything other than selected response (multiple-choice, matching, true/false). Use **Overhead A2.3,O3**—*What is "Authentic Assessment?"* to give a definition of authentic assessment.

c. Ask participants what types of assessments could be "authentic." Place **Overhead 2.3,O3**—*Authentic Assessment Can Take Many Forms* on the overhead projector, and ask participants if they agree and if their state or local assessments includes such authentic approaches.

d. An example of an authentic assessment in chemistry might be:

   i. The state chemistry or physical science curriculum instructs a teacher to teach students about acid/base reactions.

   ii. Part of acid/base reactions studies would include neutralization reactions. An acid (like HCl) or a base (like NaOH) alone could melt the flesh off the body. When the two are mixed, the neutralization reaction will always produce a salt (usually harmless to flesh) plus water (that one could potentially drink). In some cases, carbonic acid which is very unstable is produced and quickly breaks down into water and carbon dioxide. These are simple chemical reactions that occur naturally in the body.

   iii. It is the job of the teacher to take what must be taught like a chemical reaction and teach it so that students could easily interpret and answer a question similar to the one on **Overhead A2.3,O5**—*A Question Might Be*. This would be an authentic assessment that the students would never forget (they will forget memorizing chemicals like NaOH and HCl).

   iv. To explain the answer: Stomach acid is dilute HCl (hydrochloric acid) and Mom has acid indigestion, which is an acid build-up. The *Tums*™ tablet contains a base, $CaCO_3$ (calcium carbonate) to neutralize the acid. The acid and base produce salt $CaCl_2$ (calcium chloride, which is flushed out through the feces or used by the body) and H_2CO_3 (carbonic acid). She burps because the carbonic acid is unstable and breaks down into H_2O (water) and CO_2 (carbon dioxide), a gas which leaves the body as a burp. Students need to know why *Tums*™ is taken and be able to relate the chemistry behind the possible reaction. This, then, is a more familiar approach to learning about neutralization reactions.

2. (5 minutes) Tell participants that the video they are going to see illustrates the principles discussed above—assessment tasks that are related to everyday experience, and assessment that is continuous and embedded in instruction.

   Reference the information about Ms. Toliver from the video cover. Some points you might mention about this video:

   a. When teachers view the video, sometimes they comment: "Gee, if I only had the number of students she has, I could teach like her." The original number of students in Ms. Toliver's classroom was over thirty, this was too many to tape, so student names were put in a box and randomly pulled for participation.

   b. When teachers view the video sometimes they comment: "I don't have the time or energy to teach like that everyday." Listen at the beginning of the video when the students comment, "Sometimes she's the great pumpkin, etc., but most of the time she is just Ms. Toliver." Ms. Toliver does *not* conduct her lessons in this format everyday and besides, she is a 30+ year pro who has had numerous years to build her teaching style and material reserve.

   c. Most of the time, teachers will go wild over this video and will want to purchase it to show at their schools. (Ordering information is on p. 2 of this activity.)

3. (60-70 minutes) Watch and discuss the video.

   a. (5 minutes) Form small groups and pass out **Handout A2.3,H1**—*What Makes Assessment Effective in the Classroom?* Give each group one of the questions to watch for in the video and discuss when it's done.

   b. (30 minutes) Show the video.

   c. (10-15 minutes) Have groups discuss their question and think about what they do in their own classrooms.

   d. (15-20 minutes) Have each group share their question and a summary of their comments with the whole group.

Optional Extenstion—45-60 minutes to prompt more discussion

1. Ask participants to share what they think were Ms. Toliver's strongest beliefs or principles about student assessment.

2. Give participants **Handout A2.3,H2**—*Assessment Principles*. (Note that these principles were designed by a group of educators to help guide their own practice. They are meant

as discussion starters, not a model that everyone *should* use. Emphasize that it is okay not to agree with each and every point and that some principles might need clarification. Since these points are made as part of **Activity 1.12**—*Assessment Princples*, it is a good idea to do that activity first.)

Ask participants to respond to the questions on **Overhead A2.3,O6**—*A Closer Look at Assessment Principles*:

a. Which principles seemed to be illustrated by Ms. Toliver?

b. Find two of the principles that might be the easiest to implement.

c. Find one of the principles that might be the hardest to implement.

3. Give each participant two stick-it dots of the same color and one stick-it dot of a different color. Ask participants to post their dots on the chart paper graph that has sections for each of the principles. See example:

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**

4. Review the posted chart and discuss which principles the group decided were the easiest to implement and why. Discuss the principles that the group decided were the hardest to implement and why.

5. Tell participants to use this list to reflect on and guide their own growth in changing or improving on assessment practices in their own classroom.

Ms. Toliver

Purposes:

1. To discuss and examine what makes assessment effective in the classroom
2. To illustrate the term "authentic" assessment
3. To think about assessment beliefs and principles
4. To observe a teacher who models good assessment practices

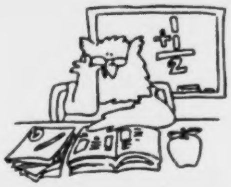
Overhead A2.3,O1

What Makes Assessment Effective in the Classroom?

- **Knowing the destination (Learning Targets)**
  - Clear beliefs or principles on what is important for students to know and be able to do
  - Crystal clear understanding of each learning target
  - Letting students "in on it"
- **Providing opportunities for practice**
  - Creating "authentic" scenarios
- **Providing good feedback on progress**
  - Oral, written, rubrics, tests
  - Student self and peer assessment

Overhead A2.3,O2

What Is "Authentic Assessment?"

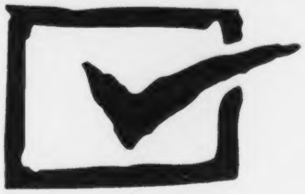


Authentic assessment tasks elicit demonstrations of knowledge and skills in ways that resemble "real life" as closely as possible.

These tasks engage students in the activity, and reflect sound instructional practice.

Overhead A2.3,O3

"Authentic" Assessments Can Include Many Formats:



The assessments can be:

- questions
- performance tasks
- oral assignments
- written assignments, etc.

Overhead A2.3,O4

A Question Might Be:

Your mother took a TUMS™ tablet last night for acid indigestion. Why? Trace the TUMS™ through her system describing the correct chemical reactions. Why did she burp?

336

Overhead A2.3,O5

A Closer Look at Assessment Principles

Look at the Assessment Principles on the Handout

a. What principles were illustrated by Ms. Toliver?

b. What are the two principles that would be the easiest for you or other teachers to implement?

c. What one principle would be the hardest for you or other teachers to implement?

Overhead A2.3,O6

What Makes Assessment Effective in the Classroom?

Group 1
Effective assessment practices include "knowing the destination (learning targets)," in other words, "knowing where you are going with learning and being able to explain why."

How did Ms. Toliver let her students know the destination (learning targets) and were her students clear about her reasons?

Group 2
Effective assessment practices provide opportunities for practice and improvement.

How did Ms. Toliver create a classroom climate that promotes student practice and improvement?

Group 3
Effective assessment practices provide opportunities for good feedback (i.e., oral, written, rubrics, grades). Research shows that students prefer oral and written feedback more than a numerical grade.

What was the major feedback mechanism used in Ms. Toliver's class? Was her method effective? Do you think her students were more concerned about their numerical grades or more curious about their learning?

Group 4
It is sometimes hard to separate good practices of assessment from instruction and curriculum. Research shows that all three must be considered together for each to be effective.

Could you separate assessment, curriculum, and instructional practices in Ms. Toliver's class? Give specific examples of each and explain why Ms. Toliver's teaching strategies seemed to work so well.

Group 5
"Authentic assessments" (real world) are assessment tasks that produce evidence of knowledge and skills in "real life" situations (or as closely as possible). These tasks not only engage students in the activity, but reflect sound instructional practices. Research shows that students respond best when conducting hands-on activities in "real life" scenarios.

After watching Ms. Toliver, give specific examples of "authentic" practices utilized in her classroom. How did her students respond to each? Are all alternative assessments "authentic?" Explain.

Group 6
What were Ms. Toliver's teaching beliefs or principles based on what she had her children do? How closely related are her principles and practices to your own?

Assessment Principles

A1

Learners need to find out often how well they are doing and teachers need to find out how successfully they are teaching. Therefore, regular assessment of student progress and achievement is part of good teaching.

A2

The main purpose of assessment is to help students learn. When students are assessed well and given feedback about their performance, they find out what they have learned successfully and what they have not. Any weaknesses can then be reduced.

A3

Assessment tasks should be designed so that most children in a group do well on most tasks. This takes the threat out of being assessed, and allows children to be motivated to learn by the regular experience of success and praise.

A4

Design/selection of assessment tasks requires a clear idea of the curriculum objectives. Children should only be assessed on knowledge, skills and attitudes their teacher has given them opportunities to develop, and each task should be well within the capabilities of most students.

A5

No single method of assessment can give information about achievement of the full range of learning objectives. Therefore a combination of different methods is vital if we are to get a balanced picture of student development.

A6

Assessment tasks must be presented in a way so that the student is perfectly clear about what is expected, and grades or marks awarded so that the student feels s/he has been fairly treated.

A7

The language of assessment must match the language of instruction. If not, then assessment produces unfair and invalid results. Children must be fluent in the language in which they are to be assessed and the level of language used must match their stage of development.

A8

The teacher's unbiased judgments are important in assessment, but students themselves can often be asked to assess their own level of achievement and the achievements of their classmates. They can be (surprisingly) accurate and honest.

A9

Assessment should focus on each student's achievements, independently of how other students are doing. Constant comparison/ competition with classmates can damage the self-esteem and self-confidence of many students.

Handout A2.3,H2

Activity 2.4

Is Less More?*

Purpose:

To consider the interrelations between classroom learning and assessment

Uses:

This is an intermediate level activity. It can be used in **Chapter 2** to illustrate integration of assessment and instruction and in **Chapter 3** to practice task and criteria development. Prerequisites might be **Activities 1.3**—*Post-it™ Notes* (definitions), **1.1**—*Changing Assessment Practices* or **1.12**—*Assessment Principles* (to look at changes in assessment and reasons for them), **1.5**—*Clapping Hands*, or **1.11**—*Assessment Standards* (to begin a discussion on quality) and **1.9**—*Going to School* or **2.5**—*How Knowledge*....(to illustrate the importance of performance criteria).

Rationale:

While there has been much talk about integrating assessment and instruction, the mechanics of how that is done can still be baffling. This activity models one approach. It involves students in self-assessment and weaves instruction and assessment together in an integrated lesson.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Flip chart and colored markers

---

* This activity is based on a classroom activity written by Ann Anderson, entitled "Mathematics in Context: Measurement, Packaging and Caring for Our Environment." Published in the March 1994 issue of *School Science and Mathematics* (Volume 94, Number 3, pp. 146-150), it is used with permission of the editors.

- **Overheads: A2.4,O1**—*Is Less More Purposes*, **A2.4,O2**—*Creating a Concept Web*, **A2.4,O3**—*Company Directive*, **A2.4,O4**—*Assessment of the Task*, **A2.4,O5**—*Some Questions*
- **Handouts: A2.4,H1**—*Is Less More?*, **A2.4,H2**—*Design Task*, **A2.4,H3**—*Problems* (Note: Before making copies of **Handout A2.4,H3**, you will need to find the population of your city/town and enter this value in the appropriate blank.)
- One per group of 3-4 people

  50-pack box of artificial sweetener
  Rulers
  Centimeter graph paper
  Scissors
  Markers and chart paper
  Calculator

- One per 2-3 groups

  Centimeter tape
  Masking tape
  White glue
  Clear tape

- Balance: electronic or with gram weights
- Purse or briefcase containing several small packages (e.g., mints, tissues, snacks), including at least one package within a package (e.g., chewing gum)

Time Required:

Part A: 35 minutes; Part B: 90 minutes

Facilitator's Notes:

Part A: (35 minutes) Concept Web

A1.(5 minutes) Introduction:

a. Facilitators introduce themselves, display **Overhead A2.4,O1**—*Is Less More Purposes*, and give a brief overview of session:

We will experience an example of a mathematics and science lesson that might be typical of today's standards-based classroom and consider the implications of such a lesson for student assessment.

b. Divide participants into groups of 3-4.

c. Give each person **Handout A2.4.H1—*Is Less More?*** Read the lesson objectives as participants follow along. Then tell participants:

Try to be aware during the activity of when you think you personally have met each of the lesson objectives. Feel free to make notes on the handout to help you remember the context in which the objective was met.

A2.(10 minutes) Hook

a. Display packaged items at appropriate times while saying:

I was cleaning out my purse (or briefcase) the other day and was really surprised at all the different packages that I carry around. And then I started thinking about how some things are even packages within packages.

Demonstrate by removing stick of gum from outer gum packaging.

When I went grocery shopping later that day, I found that I was suddenly very aware of all the packages within packages that I purchase.

Hold up a box of artificial sweetener and ask:

Do any of you purchase this product or one like it? Here we have 50 little packets holding the sugar substitute and then packed inside a larger box.

b. Give each small group a box of the artificial sweetener packages. **Caution them not to open the boxes yet.** Ask groups to shake the box up and down and ask what they notice. Elicit from participants the idea that the box is not full. Then ask:

Do you think we could make the box smaller and still have it hold 50 packets of sugar substitute

Participants should say "yes" or "probably." Then ask:

Why might we want to use a smaller box?

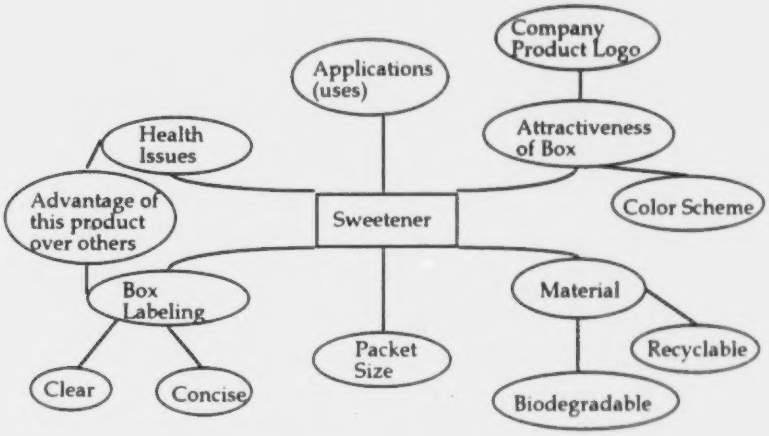
Take a few responses from participants. Responses may include environmental, marketing, or economic concerns. If you want to stress science as a focus, and environmental concerns are not raised, you may want to suggest them. Then state:

Given that there may be some good reasons for making a smaller box, I wonder why the company has chosen to make the box this size.

A3. (10 minutes) Concept Web

a. Use **Overhead A2.4,O2**—*Creating A Concept Web* to guide groups in exploring factors that may have influenced the design of the box. Have each group draw the box in the center of the chart paper, brainstorm factors that might have influenced the design of this package, and draw a concept web depicting the influences.

(Note: A sample concept web generated in a recent workshop follows.)


Company Product Logo
Applications (uses)
Health Issues
Attractiveness of Box
Advantage of this product over others
Sweetener
Color Scheme
Box Labeling
Material
Recyclable
Packet Size
Clear
Concise
Biodegradable

A4.(10 minutes) Large Group Sharing:

Call on one person from each group to briefly describe the group's concept web. Allow about 2 minutes per group and ask for comments or questions.

Part B: (90 minutes) Design Task

B1. (10 minutes) Company Directive

a. Tell participants that you understand that the artificial sweetener company is interested in designing a smaller box to hold the 50 packets. Display the Company Directive (**Overhead A2.4,O3**). Explain that each group will be given the opportunity to meet the Company Directive. Groups are going to take on the role of designers and attempt to help the company with this task by trying to design a smaller box.

b. Ask participants to look at their box again. Pose the following question:

If I asked you to design a box that used less material and to prove mathematically that it used less, what would you have to know or be able to do?

Capture their responses on a blank transparency. Be sure to elicit from participants that one thing they would have to know is how to find out how much paperboard there is in the original box and in the new box. Give groups **5 minutes** to discuss how they would determine how much material is in the box. Call on one person to share ideas from each group. Be certain participants know what materials are available (e.g., rulers, tape, glue, the triple beam balance). Important: Check to see that everyone understands the different ideas before going further.

c. The Design Task: Distribute (one per group) **Handout A2.4,H2**—*Design Task* and explain the Design Task that addresses the Company Directive. Participants will design (and if desired, use the existing box to build) a smaller box that can still hold all 50 packets of the sweetener. They will record all measurements and computations, and compute the percentage reduction in paperboard of the new box. DO NOT ALLOW PARTICIPANTS TO BEGIN THE DESIGN TASK YET!

B2. (10 minutes) Assessing the Task

a. Use **Overhead A2.4,O4**—*Assessment of the Task* to tell participants that before they work on the design task, they need to consider how they might assess successful completion of the task. Remember that the task is designed to practice working cooperatively as well as to practice problem solving and application of knowledge.

b. Allow 8-10 minutes for groups to develop criteria/strategies for assessing the quality of performance on the design task, including the actual package design. Stress the importance of writing value-neutral statements that encompass each important objective of the activity and can easily be assessed. Be sure each group has included a strategy for assessing mathematical and/or science skills, concepts, and group skills. Make sure chart papers are easy to read and visibly posted.

B3. (30 minutes) The Design Task

a. Tell participants they will have **25 minutes** to re-design the box and complete the design task. Point out the variety of materials available on the supply table for their use. **Tell participants they should keep in mind the assessment strategies/criteria that are posted around the room as they complete the design task.**

b. Pass out the problems on **Handout A2.4,H3**—*Problems*. The problems cannot be completed until they know the percentage reduction in paperboard of the new box.

As participants work, the facilitator should move around the room with notepad in hand. Refer to each group's posted criteria and jot down notes about your perception of the extent to which each group is meeting their criteria. Be careful not to compare groups.

c. Ask groups to prepare a display of their completed boxes, problems, and summary paragraph at their tables.

B4. (5-10 minutes) Round Robin Sharing. Have groups rotate from table to table to quickly examine each group's finished product. Allow no more than 1 minute per station. (Or, see Optional Extension.)

B5. (10 minutes) Assessing Student Performance and Lesson Objectives

Point out that performance on the activity was assessed using a variety of strategies for a variety of purposes. On flip chart sheets, have small groups prepare a list of different strategies that were used, along with the purpose of each. You may need to stimulate the discussion further by providing any of the following examples:

a. To assess individuals' attainment of lesson objectives: self-reflection, confirmation (or not) by peers, teacher observation during group work

b. To assess group performance: group reflection on how it met its own criteria, production of a product that met specifications, self or teacher examination of problem sheets, and summary paragraph

You may also wish to say that in an ideal classroom setting, the facilitator would later collect and review each group's written calculations, summary paragraph, and problems. S/he would then meet with each group for a consultation to compare the instructor's assessment of the task with the group's assessment. Together, they would then identify areas for additional study and develop suitable activities for building strengths and enhancing learning of specific skills and knowledge.

B6. (15 minutes) Closure

Referring to the list of assessment strategies generated above, use **Overhead A2.4,O5**—*Some Questions* to facilitate a discussion about the following questions:

Are these assessment ideas and strategies different from traditional assessment approaches? If so, how are they different?

What are practical realities of using these kinds of assessment strategies in the classroom?

What are some implications for teachers, parents, course grade, etc.?

Close by acknowledging that there are a variety of assessment approaches that can be integrated into instructional lessons. Each provides a different kind of information about the students. Traditional assessments allow limited approaches to assessing students. By incorporating a wide variety of assessment strategies into classroom teaching, different student and teacher learning and teaching styles can be accommodated.

Optional Extension—60 minutes

After groups have viewed each other's displays, have them select one of their criteria and generate an expanded description of what might be considered an excellent response. After sharing these with the large group, facilitate a discussion about how samples of student work allow refinement of the criteria.

Is Less More?

Purposes:

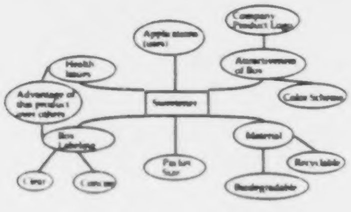
1. To stimulate discussion about the relationship of classroom learning and assessment

2. To experience an integrated, interactive standards-based mathematics and science lesson and consider the implications of such a lesson for student assessment

Overhead A2.4,O1

Creating a Concept Web

1. Sketch the box in the center of the chart paper.
2. Brainstorm different factors that might have influenced the design of the box.
3. Use lines to connect the box with words, phrases, or sketches derived from your list of factors.
4. The resulting diagram is a concept web that shows the different influences on the design of the box.


Company Product Logo
Health Issues
Attractiveness of Box
Advantage of this product over others
Color Scheme
Box Labeling
Material
Recyclable
Biodegradable

Overhead A2.4,O2

BEST COPY AVAILABLE

Company Directive

With increased pressure from environmentalists, the company has decided to solicit assistance in re-designing their box. Groups are invited to:

Submit a design and/or model for a box that uses less of the same kind of material.

Submit mathematical proof that the new design uses less material than the current design.

Overhead A2.4,O3

349

Assessment of the Task

How might the Design Task be assessed?

1. Develop criteria for assessing the quality of performance on the task. Include criteria for assessing mathematical/science skills, conceptual knowledge, and group skills.
2. Write these on flip chart sheets (big) and post the sheets next to your work station.
3. Write a summary paragraph that states how well your group thinks it met the criteria it established.

Overhead A2.4,O4

Some Questions

Are these assessment ideas and strategies different from traditional assessment approaches? If so, how are they different?

What are practical realities of using these kinds of assessment strategies in the classroom?

What are some implications for teachers, parents, course grade, etc.?

How does the use of tasks like this one illustrate "less is more" in teaching and learning?

What do we give up when we engage students in integrated tasks with embedded assessment? What do we gain?

Overhead A2.4,O5

IS LESS MORE?

"Paper production has grown rapidly throughout the twentieth century, increasing its output twenty fold since 1913. Rising literacy around the world accounts for some of this growth, but the largest share of paper and paperboard is used in industrial countries as **boxes and other forms of packaging."** (From *State of the World 1994* by Lester R. Brown, et al.)

LESSON OBJECTIVES

During this lesson, participants will:

1. Increase awareness of issues related to packaging consumer products
2. Explore multiple designs for a box of sugar substitute
3. Apply measurement and computation skills in a real world setting
4. Work cooperatively to produce a group product
5. Identify a variety of assessment strategies

Handout A2.4,H1

Design Task

Work cooperatively to design (and build) a smaller box that still contains all 50 packets of sugar substitute.

Record all of your measurements (with units) and show computations.

Compute the percentage reduction in paperboard from producing your new box instead of the original box.

Be sure all members of your group understand the group's work well enough to explain it to others.

Handout A2.4,H2

PROBLEMS

Consider the original source of the paperboard in a non-recycled box: a tree. On the average, a 70 foot (21 m) tall tree that is 10 inches (25 cm) in diameter provides about 200 pounds (90 kg) of paper.

PROBLEM #1: If one artificial sweetener box has a mass of 20 grams, how many boxes can be produced from one tree? (Hint: 1000 grams = 1 kilogram)

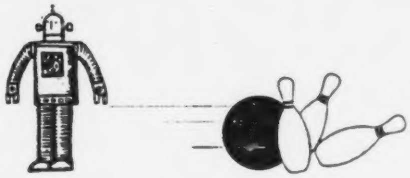
PROBLEM #2: The population of __________________ is approximately __________________ people. Assume that 10% of consumers purchase this brand of artificial sweetener at a rate of 3 boxes per month. How many trees must be cut in one year to produce these boxes to satisfy your city's consumers?

PROBLEM #3: How many trees could be saved in one year if these consumers purchased your new smaller box of sweetener?

Handout A2.4,H3

Activity 2.5

How Knowledge of Performance Criteria Affects Performance— Robotic Arm or Bowling Scores



Purposes:

1. To illustrate the importance of performance criteria for helping students understand the requirements of an assignment or task
2. To illustrate what can happen when criteria for judging success are not clear
3. To illustrate that performance criteria are really part of the task description
4. To illustrate the required match between performance criteria and tasks

Uses:

This is an intermediate level activity. It can be used in **Chapter 2** to illustrate the importance of criteria as a tool for communication and learning in the classroom. It could also be used as a follow-up to activity **1.5**—*Clapping Hands* to remind folks of the importance of up-front criteria to give to students. In **Chapter 3** it can be used to demonstrate how performance criteria help define the task. A prerequisite would be **Activity 1.3**—*Post-it™ Notes* (definitions).

Rationale:

Many times teachers don't realize the extent to which a task might be confusing or misleading for students. This activity is designed to illustrate how, without clearly stated criteria, students might not approach an activity in the way intended by the teacher. Stated another way, having clearly stated criteria for success on an activity assists students in knowing what to do.

Note: Although the *Robot Arm* handouts (**A2.5,H1** and **A2.5,H2**) are a science task and *Bowling* handouts (**A2.5,H3** and **A2.5,H4**) are a mathematics task, this activity emphasizes both student process skills and student knowledge of science or mathematics content, so it can be used in many contexts.

Materials:

- Overhead projector, screen, blank transparencies and pens or flip chart pens
- **Overheads A2.5,O1**—*How Knowledge...Purpose,* and **A2.5,O2**—*Discussion Questions*
- Copies of either **Handout A2.5,H1**—*Tube and Tendon Robotic Arm* and **Handout A2.5,H2**—*Tube and Tendon Robotic Arm Criteria*, or **Handout A2.5,H3**—*Bowling* and **Handout A2.5,H4**—*Bowling Criteria*
- For *Robot Arm* you will need (for each 6 people): scissors, string, straight pins, straws, and masking tape

Time Required:

30-45 minutes

Facilitator's Notes:

1. (5 minutes) Introduction

   Distribute handouts for either *Tube and Tendon Robotic Arm* (**A2.5,H1**) or *Bowling* (**A2.5,H3**). For *Robotic Arm*, divide the participants into groups of about 6 and give one set of materials to each group. For *Bowling*, you may wish to have a small volunteer group of about 6 participants solve the problem aloud in front of other participants. (This is a "fishbowl" technique.)

2. (5-10 minutes) Do the task.

   a. Start groups working on the assigned task (without telling them what the criteria are).

   b. After about 5 minutes, stop the groups and ask them what they think the criteria are for success. For *Robot Arm*, participants usually judge success by the extent to which the robotic arm "works." For *Bowling*, participants usually judge success by how close the group feels they are to getting the right answer.

3. (5 minutes) Tell participants the teacher's intended criteria for success on the task using **Handouts A2.5,H2**—*Tube and Tendon Robotic Arm Criteria* or **A2.5,H4**—*Bowling Criteria*. These handouts illustrate the difference between what participants originally thought were the criteria for success and other possibilities for criteria.

4. (5-10 minutes) Ask groups to alter their course (as needed) to match the criteria.

5. (5-10 minutes) Discuss the questions on **Overhead A5.2,O2**—*Discussion Questions*.

6. (5-10 minutes) Summarize with **Overhead A2.5,O1**—*Purposes.*

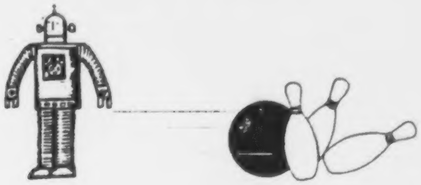
How Knowledge of Performance Criteria Affects Performance

Purposes:

1. To illustrate the importance of performance criteria for helping students understand the requirements of an assignment
2. To illustrate what can happen when criteria for judging success are not clear
3. To illustrate that performance criteria are really part of the task description
4. To illustrate the required match between performance criteria and tasks

Overhead A2.5,O1

Discussion Questions Robotic Arm or Bowling Scores



1. How did your actions change once you knew the criteria?

2. How would you have felt if you had gotten the criteria after you completed the task?

3. What are the advantages of having the criteria "up front"?

4. What work do your students do now for which it would be prudent to give them criteria in advance?

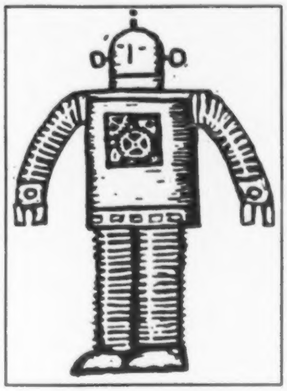
Overhead A2.5,O2

358

Tube and Tendon Robotic Arm Task

Materials:

- Straws
- String
- Masking tape
- Scissors
- Straight pins



The Challenge:

With your group, design a robotic arm that can bend in two or more places. Show in your design how you will cause the arm to bend and to flex straight again.

The Steps:

- Design your arm on paper (prototype).
- Engineer a prototype, using your materials.
- Test the prototype, and record how well it works.
- Revise your design based on results of your test.
- Ask, "Is there any device in the real world like yours?"
- Ask, "How could somebody use this?"

Activity prompted by concepts in Susan Dunn and Rob Larson, Design Technology: Children's Engineering. New York: The Falmer Press, 1990.

Handout A2.5,H1

Tube & Tendon Robotic Arm Criteria

What the students PERCEIVED to be the criteria:

To design a robotic arm that would actually work—that would flex in two or more places, then extend again

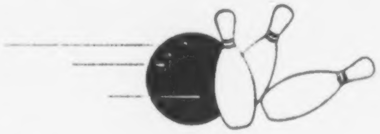
The teacher's ACTUAL criteria:

- Effective, cooperative group work: EVERY member participates
- Careful design of prototype
- Perseverance: Willingness to try again (and again)
- Inventiveness in redesigning the original prototype

360

Handout A2.5,H2

Bowling



The tables below show some bowling scores. A higher number indicates a better game.

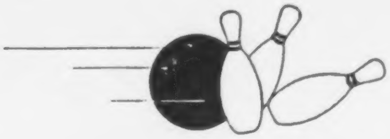
Dave's Scores		**Bill's Scores**	
152	138	210	105
138	141	105	105
141	141	118	118
144	144	131	131
141	152	105	210
158	158	215	215

Both Dave and Bill are trying out for the bowling team. As the coach, you must decide which one should be picked to join the team.

Examine the data and use it to justify your choice.
Explain to Dave and Bill how you came to your decision.
Use charts or drawings if it helps you to explain more clearly.

Handout A2.5,H3

Bowling Criteria



What the students perceived the criteria to be:

How close are we to solving the problem?

The teacher's ACTUAL criteria:

- Demonstrate understanding of central tendency and spread of data
- Effective, cooperative group work—EVERY member participates
- Perseverance—willingness to keep trying
- Problem solving
- Communication in math

Handout A2.5,H4

362

Chapter 3
Designing High Quality Assessments

What's in This Chapter?

While **Chapters 1** and **2** strive to help users develop an overall vision of the role of assessment in instruction and explain how purposes affect assessment design, **Chapter 3** begins a detailed examination of design options and related quality considerations. What do some of the current assessments look like? When should various design options be used? How do assessment purposes relate to design options? What do high quality alternative assessments look like? These questions, relevant to both large-scale and classroom assessment, are addressed in this chapter.

Chapter Goals

1. Illustrate design options for alternative assessments
2. Expand expertise in developing alternative assessments
3. Present guidelines for assessing the quality and appropriateness of alternative assessments for particular purposes and contexts
4. Provide experience in choosing high quality assessments
5. Examine issues of equity, fairness, bias, and unintended consequences of assessment

Chapter Content

A. Readings

Background Information...Page 4

This section sets up the rest of the chapter by specifying the two required parts of an alternative assessment (tasks and criteria) and the need for quality.

Performance Tasks—Design Options and Quality Considerations.......Page 7

The first required element of an alternative assessment are tasks—activities for students to do. This section describes design alternatives for tasks, provides advice on when to use various designs, and discusses quality issues that must be addressed when designing tasks.

Performance Criteria—Design Options and Quality Considerations ....Page 20

The second required part of an alternative assessment are criteria that specify the basis on which the quality of student performance on the task(s) will be judged. This section describes design alternatives for criteria, provides advice on when to use various designs, and discusses quality issues that must be addressed when designing criteria.

Quality—The Rest of the Story ....Page 32

This section discusses quality issues that go beyond those for tasks and criteria—adequate content and skill coverage, fairness and bias, consequences of assessment, and cost/efficiency.

Conclusion ....Page 39

Summary Rating Form ....Page 40

The quality issues described in this chapter are summarized in a rating form that can be used to review alternative assessments for quality.

B. Activities in This Chapter

Activity 3.1 *Performance Tasks—Keys to Success*

This advanced level activity illustrates the dimensions along which performance tasks differ and discusses the advantages and disadvantages of different approaches. It uses example assessments from **Appendix A**. Time: 60 to 90 minutes

Activity 3.2 *Spectrum of Assessment Activity*

Part A of this activity asks participants to review Bloom's Taxonomy and develop assessment questions and tasks that tap different levels. In Part B, participants review a variety of short, related assessment tasks and rank them according to the type of skills and knowledge each might elicit from students. The goal is to design tasks that really assess desired learning targets. This is an advanced level activity. Time: Part A, 90 minutes to 2 hours; Part B, 90 minutes

Activity 3.3 *Performance Criteria—Keys to Success*

This advanced level activity illustrates the characteristics of sound performance criteria. It uses assessment samples from **Appendix A**. Time: 60-90 minutes

Activity 3.4 *Assessing Learning: The Student's Toolbox*

This intermediate activity illustrates the relationship between different ways to design assessment tasks and the student learning to be assessed. Time: 20-30 minutes (variation takes 45-60 minutes)

Activity 3.5 *Performance Tasks and Criteria: A Mini-Development Activity*

In this advanced level activity, participants discuss how good assessment equals good instruction and the issues in using and designing performance assessments. Participants use assessment samples from **Appendix A**. Time: 2-3 hours

Activity 3.6 *How to Critique an Assessment*

In this advanced level activity, participants practice aligning assessment and instruction and evaluating a performance assessment for quality. Time: 1-2 hours

Activity 3.7 *Chickens and Pigs: Language and Assessment*

This introductory level activity emphasizes the critical role that language plays in effective and equitable assessments. Time: 10-20 minutes

Activity 3.8 *Questions About Culture and Assessment*

This introductory level activity increases awareness of the relationship between culture and assessment by demonstrating how cultural factors can affect student ability to show what they know and can do. Time: 90 minutes

Activity 3.9 *Tagalog Math Problem*

This introductory level activity illustrates how lack of knowledge of a language can affect performance on skills that have nothing to do with language understanding (like math). Time: 20-30 minutes

Designing High Quality Assessments

Readings

Tasks and Criteria–Overview

Tasks and Performance Criteria–You Need Both to Have an Assessment

We've looked at hundreds of alternative assessments in order to discover *how* they are being designed and *why* they are being designed that way. Since we wanted to get lots of ideas, we used a fairly broad definition of alternative assessment—see the box. We did, however, have one requirement—to count as an "assessment" the instrument, technique, or procedure had to have two parts:

1. Specification of *tasks*. The *task* is the activity or assignment the students are asked to do. Student responses to the task are what is assessed. *Tasks* can be **separate** exercises or problems given to students at the end of instruction just to assess achievement. But, **tasks can also be regular classroom activities** during which teachers collect information (for example, noting

🗀 Definition

Assessment terms are used differently by different people. Here is how we've agreed to define "alternative assessment."

Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.

Some people also call these types of assessments "constructed response."

366

communication ability during class discussions). But, in either case, for us to include the assessment, the task had to be described.

2. *Performance criteria*—a specified method of evaluating student performance on the tasks. Methods could include: assignment of points for the "correctness" of responses, checklists of response features, and rating scales requiring professional judgment.

It is not assessment if these two components are not present.

A common tendency for those just starting out with alternative assessment is to focus on the task to the exclusion of the criteria. Many think of assessment as only the tasks given to students to do. This is a holdover from past experience with tests. On multiple-choice tests, for example, since scoring is right/wrong, the problem or question given the students *is* the major consideration. If the student gets the question correct, it is assumed that he or she knows whatever content or skill the question covered.

However, with alternative assessments the manner of scoring—the performance criteria—is as important as the tasks. The criteria actually define what is assessed more than the task does. Consider the situation in which the performance criteria for a math task emphasize conceptual understanding. It doesn't matter if the task requires problem solving, making a presentation, collaboration, or anything else—only conceptual understanding is assessed.

Related Toolkit98 Chapters and Activities

Activity 1.5—*Clapping Hands* discusses quality considerations for tasks and criteria as two important parts of an alternative assessment.

Activity 1.9—*Going to School* illustrates the importance of performance criteria when making consistent judgments about students.

Activity 2.5—*How Knowledge of Performance Criteria Affects Performance* illustrates the importance of criteria in helping to define the task.

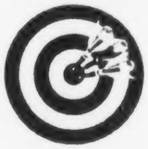
Or, consider an assessment in which students have to estimate the number of beans in a bucket using a cup, ruler and scale. What this task actually assesses depends more on the criteria used to judge success than on the task itself. If it were scored right/wrong (with the student getting it "right" if his or her estimate were close to the real number of beans) then one might assume that the test measured ability to estimate. However, the *actual* criteria for success in this case focus on problem solving and communication. A high score is given not for getting the right answer, but for using good problem-solving strategies and adequately communicating what was done. The task of estimating is just the context for assessing problem solving and communication. What is assessed is specified by the criteria, not the task.

Yet, it is not uncommon for those who are just beginning to develop or choose alternative assessments to attend only to development of the *tasks*. One might hear, for example, a teacher saying, "I do alternative assessment; I have students write journals," or "I do alternative assessment; I have students do computer simulations." But, the question remains, *What do you look for in those journal entries or in the response to the simulation that will tell you whether you are accomplishing your learning goals for students? What will you assess? How will you know when performance is good enough?* These questions are answered by the performance criteria, not by the task.

Alternative Assessment Doesn't Automatically Ensure High Quality Assessment

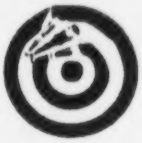
Just because an assessment looks "interesting" or "authentic" does not mean it is high quality. It's perfectly possible to do a poor performance assessment, poorly conceive and execute a portfolio system, or produce assessment activities that do not address valued student learning targets. Two important terms used when discussing the quality of assessments are "validity" and "reliability."

- **Validity.** A valid assessment measures what it is intended to measure and permits the user to draw accurate and fair conclusions about students' performance. For example, if an assessment is intended to be used to draw conclusions about an individual student's mastery of writing skills, the assessment is valid to the extent that it really does allow one to draw that conclusion. Or, if an assessment is intended to diagnose student strengths and needs in reading, the assessment is valid to the extent that it allows one to do that. An assessment valid for one use may not be valid for another use. An assessment that is valid hits the right target.



  Now, this is very easy to say, but how does one actually determine whether an assessment hits its mark? Against what do you compare it? Well, one way is to look at the content and see if it looks like it covers the right stuff or represents things students have to do in real life. Assessment folks call this "content validity," "ecological validity," or "face validity" (really!). Another way is to see whether students who do well on other, similar tasks also do well on this task. Assessment folks call this "concurrent validity" or "predictive validity." A third way is to try and determine whether the assessment corresponds to theories about learning or cognition. For example, does the task distinguish between experts and novices? Assessment folks call this type of validity "construct validity." A fourth way is to try and determine the unintended consequences of an assessment. If good things happen as the result of an assessment, it has "consequential validity."

- **Reliability** refers to how repeatable the scores are. Would the conclusions drawn about student ability be the same regardless of when in the school day the assessment was given, which form of the assessment the student received, or which rater happened to score the work? Do equal scores indicate equivalent performance across students, raters, and days? How comfortable would you feel, if the same student work received vastly different scores from two different raters, or from the same rater on two different days? Obviously, if an assessment is not reliable, it cannot be valid.



The target pictured above illustrates validity *and* reliability—each individual arrow hits the same place regardless of who fired it, when it was fired or which bow was used. The target to the right illustrates an assessment that is reliable, but not valid—the arrows consistently hit the same place, just not the *right* place.

Although the formal terms "reliability" and "validity" are not used very much in this chapter, they're really what this chapter is all about—thinking about the reliability and validity concerns of alternative assessment. Familiarity with these issues is important for both classroom or large-scale assessment. Assuring the validity of assessment is easier said than done. It requires consideration of a number of assessment features, which are described throughout this chapter.

Chapter Organization

First we'll consider design options and quality considerations for performance tasks. Then we'll consider design options and quality considerations for performance criteria. Finally we'll review quality considerations that span tasks, criteria, and the context in which the assessment will occur. Finally, we'll summarize it all in a rating form that can be used to evaluate the quality of alternative assessments.

Performance Tasks—Design Options and Quality Considerations

Task Design Options

Performance *tasks* are the exercises or assignments given students to do; it is student performance on the *task* that is assessed. There are differences in how tasks are designed. These differences reflect one's belief about education and philosophy of instruction. Therefore, it is useful to discuss some of the variations, so that the reader can relate the possibilities to his or her own instructional philosophy and goals for students.

Open-Response vs. Open-Ended Tasks . In *Toolkit98* the term "open-response" refers to problems or tasks for which the student generates an answer, but there is still only one right, or a few best, answers. So, even though tasks may be complicated and require equipment or manipulatives, and even though we might assess student performance on the problem-solving process as well as getting the right answer, there is still only one right, or a few best, answers. All the examples in **Figure 3.1** are open-response questions. (Additional examples are provided in **Appendix A**—*Alternative Assessment Sampler*.)

Figure 3.1
Examples of Open-Response Tasks
(Numbered samples can be found in **Appendix A**—*Sampler.)*

A. 145 + 28 + 3767 + 90405 = _______________ (Arithmetic Problem)

B. "Estimate the number of beans in the bucket." (Students are given a large bucket of beans, magic markers, a tray and a small cup—*Bucket of Beans*; see **Sample A.34**—*Performance Assessment in Mathematics, Alberta*.)

C. "Draw a picture of what the weather looks like today in the 'window' on the top half of the page. Make your picture as complete as you can." (See **Sample A.7**—*Weathercaster's Helper*.)

D. "Examine the items in the garbage can. Sort the items into the following categories: paper, metal, food waste, plastic, yard waste, glass, other. Count and record the number of items in each category. List the categories in order from least number of items to greatest number of items." (The purpose of the exercise is to see whether students can classify different types of solid waste. This is open-response because *all* students are given a garbage can with the *same* trash in it and there is only one "right" way to sort it. See **Sample A.16**—*Solid Waste*.)

E. "Do sow bugs prefer it where it is light or where it is dark?" (Students are given experimental equipment. See **Sample A.23**—*Sow Bugs*.)

F. "Your task is to determine the location of specific features of the simulated ocean floor...Produce a graph which represents a profile of the ocean floor." (Students are given a black box with a simulated ocean floor and scale, and a height finder with a sliding ring. See **Sample A.32**—*Mapping the Blue Part*.)

In *Toolkit98* the term **open-ended** refers to problems or tasks for which there is no single correct answer; students could generate any number of acceptable responses. Consider the tasks in **Figure 3.2**.

Figure 3.2
Examples of Open-Ended Tasks

(Numbered samples can be found in **Appendix A**—*Sampler.)*

G. "Your class will be getting a 30-gallon aquarium. The class will have $25.00 to spend on fish. You will plan which fish to buy. Use the *Choosing Fish for Your Aquarium* brochure to help you choose the fish. The brochure tells you things you must know about the size of the fish, how much they cost and their special needs. Choose as many different kinds of fish as you can. Then write a letter to me explaining which fish you choose. In your letter tell me how many of each kind of fish to buy, give the reasons you chose those fish, and show that you are not overspending and that the fish will not be too crowded in the aquarium." (*Aquarium*; see **Sample A.15**—*Aquarium Problem.*)

H. Students produce a mechanized vehicle that produces at least two simultaneous motions in different directions to accomplish a set of actions. For example, a dump truck or an oscillating fan. (*Mechanized Vehicle*; see **Sample A.44**—*High School Physics.*)

I. "Develop and write a tall tale or fable." (*Tall Tale*; see **Sample A.9**—*Tall Tales & Fables.)*

J. "What could you do to reduce the amount of solid waste in this garbage can?" (This is the final part to example D shown above, **Sample A.16**—*Solid Waste.*)

K. "What do you think this might be the graph of? Put names and numbers on the graph to show what you mean." (Students are given a graph with no labels—*Graph Problem.* See **Sample A.18**—*Mathematics Problem Solving, Oregon.*)

L. Students use primary and secondary source materials to explore an historical issue. After reading and discussing documents, groups decide on a positions to take and individuals write a position paper. (*History Paper*; **see Sample A.38**—*Historical Investigations, California.*)

M. Students prepare a science portfolio in which they select one problem-solving investigation, one piece of work that demonstrates creative expression, and one piece of work that demonstrates growth through writing. They explain why each piece was chosen. (See **Sample A.42**—*Science Portfolio.*)

We have found many examples of both types of tasks. One approach is not necessarily better than another. Open-response tasks seem to be associated with the attempt to measure knowledge of specific pieces of information or the ability to apply specific concepts to new situations, while open-ended tasks seem to be associated with the attempt to measure problem solving, group process, critical thinking, or science process skills. It stands to reason that tasks having many possible correct responses might be more useful in assessing skills like decision making.

This doesn't mean that's what **you** should do. There are also assessments in which students are given open-response tasks, and student work is scored for problem solving as well as the correct answer. What you *should* do is start with a clear idea of what you want to measure and a well-articulated instructional philosophy and approach (see **Chapter 2**), then choose or design tasks that will fit your instructional philosophy and elicit the desired behavior on the part of students.

How Students Perform Tasks. There seem to be two major ways that students perform tasks—with and without the use of manipulatives or equipment. In *Toolkit98* the use of manipulatives or equipment is called "hands-on." Hands-on tasks contrast with assessments that are done only with paper and pencil, in which students may manipulate symbols, draw pictures, or do calculations, but all on paper. See **Figure 3.3** for examples of each type.

Figure 3.3
Examples of Hands-On and Paper and Pencil Tasks

Hands-On	All Paper and Pencil
B—Bucket of Beans D—Solid Waste E—Sow Bugs F—Mapping the Blue Part H—Mechanized Vehicle	A—Arithmetic Problem C—Weathercaster's Helper G—Aquarium I—Tall Tale K—Graph Problem

Both approaches (with and without the use of equipment and manipulatives) have been used to assess all kinds of outcomes: knowledge, skills, conceptual understanding, and "big" outcomes such as problem solving, communication, group process skills, and critical thinking. Again, one procedure is not necessarily better than another. Each technique has advantages and disadvantages. Users need to balance the following factors:

- **Purpose for the assessment.** If the purpose is to quickly gather information on many students (large-scale assessment), it's more efficient if the tasks are paper and pencil.
- **The skills measured.** Some skills, such as science process skills, require hands-on materials.
- **Developmental level of students.** Younger students need more hands-on activities.
- **Instructional philosophy.** If instruction emphasizes manipulatives, it would be a good idea to also use them on the assessment. On the other hand, we must not fall into the trap of blindly using hands-on tasks because kids like them. There has to be a good pedagogical reason to use a hands-on task.
- **What the assessment communicates** to teachers and students about what is important. If we say that teachers should do hands-on activities, but assessment is always paper and pencil, it sends a message about what is important.
- **Practical necessity.** Sometimes the logistics of a hands-on assessment are too overwhelming to contemplate.

Response Mode. Assessments we've looked at also differ in how students are asked to respond. Most student responses are written, some are oral, and some require a physical response or creation of a product (see **Figure 3.4**). Note that responses could be various, such as example H, in which students produce a product and present an oral report about it.

Figure 3.4
Response Mode of Tasks

Written Response	Oral Response	Physical Response/ Product
A—Arithmetic Problem B—Bucket of Beans C—Weathercaster's Helper D—Solid Waste F—Mapping the Blue Part G—Aquarium I—Tall Tale K—Graph Problem L—History Paper M—Science Portfolio	H—Mechanized Vehicle L—History Paper	E—Sow Bugs H—Mechanized Vehicle

Written responses are more practical and produce a permanent record. That's why they're used so much. Producing products also is fairly permanent although bulkier and might be a better way to assess some outcomes—for example, if one wants to really see if a student can *do* something and not just talk about doing it. Oral responses might be more appropriate for younger students, students with poor writing skills, and students for whom English is a second language. *Care must be taken that a student's writing ability does not mask the actual student skill being assessed.*

Length/Complexity. We're beginning to hear about "three P's"—on-demand Performance assessment, Projects, and Portfolios. *On-demand performance assessments* are relatively short tasks (spanning one to three class periods) that students do at more or less the same time in more or less the same manner. (That is, they tend to be more standardized.) Some authors distinguish "short answer" on-demand tasks from "on-demand performance assessments." Between the two, on-demand performance assessments tend to require longer responses.

Projects are longer term activities that can take the place of other instructional activities and are embedded in classroom work. Projects used for assessment are frequently indistinguishable from projects used for instruction. In fact, projects used for assessment should model good instruction. A good discussion of this can be found in Gong et al. (1992).

Portfolios are collections of student work that are assembled over fairly long periods of time and are even more instructionally embedded. We use this definition for portfolios:

> *A portfolio is a purposeful, integrated collection of student work showing effort, progress, or degree of proficiency. The assessment potential and instruction usefulness of portfolios are enhanced by clear learning targets that focus content, criteria that define quality, and student involvement to make learning active.*

📖 **Reference Box**

Brian Gong, R. Venezky, and D. Mioduser, 1992. "Instructional Assessments: Level for Systemic Change in Science Education Classrooms," *Journal of Science Education and Technology*, Vol. 1, pp. 157-176.

Samples in portfolios are frequently chosen by students to demonstrate what they know and can do. Thus, portfolios tend to look very individual. Standardization, when it occurs, tends to concentrate on the categories of items students must submit for their portfolios (*clear learning targets* in the definition above) and in the *criteria used to define quality*, rather than mandating particular work or performances for inclusion.

See **Figure 3.5** for examples of on-demand performance assessments, projects, and portfolios.

Figure 3.5
Length and Complexity of Tasks

Short Answer	On-Demand	Project	Portfolio
A—Arithmetic Problem	B—Bucket of Beans C—Weathercaster's Helper D—Solid Waste E—Sow Bugs F—Mapping the Blue Part G—Aquarium I—Tall Tale	H—Mechanized Vehicle L—History Paper	M—Science Portfolio

Figure 3.6
Group Work

Individual Work	Group Work	Both
A—Arithmetic Problem B—Bucket of Beans C—Weathercaster's Helper D—Solid Waste E—Sow Bugs G—Aquarium I—Tall Tale K—Graph Problem M—Science Portfolio	F—Mapping the Blue Part H—Mechanized Vehicle	L—History Paper

Which options to select depends on what learning targets are to be assessed and the purposes for assessing them. For example, if one needs to make high-stakes decisions about individual students, one needs to gather sufficient evidence to be sure that the estimation of student ability is valid. Since some of the skills on which we want to judge competence require complex behaviors, a portfolio might be called for. A portfolio allows the collection of multiple examples of complex performances over time. No single on-demand performance assessment allows this. A perfect example of the need for portfolios is the current Certificate of Mastery reform effort in several states, with students having to demonstrate mastery of complex abilities such as writing, math problem solving and oral presentations to receive an endorsement on their diploma. Adequate documentation of student ability requires several samples of work from different content areas and assignments. Portfolios seem a logical answer.

Our point here is that task type should not be chosen just because it's faddish or sounds like a good idea. Tasks must be crafted to accomplish what is desired. If the purpose is a quick, overall estimate of how a group of students is doing on math problem solving, for example, an on-demand performance assessment of five problems randomly distributed among students (so that any single student only gets one problem) might be the best choice. If the purpose is to make high-stakes decisions about individual students, a combination of approaches might be necessary.

⊂ Related *Toolkit98* Chapters and Activities

Activity 3.1—*Performance Tasks—Keys to Success*, **Activity 3.2**—*Spectrum of Assessment Activity*, and **Activity 3.4**—*Assessing Learning: The Student's Toolbox*, and **3.5**—*Performance Tasks and Criteria...*, give participants the opportunity to look at different task types and discuss their advantages and disadvantages based on the considerations outlined in this section.

Activity 3.5—*Performance Tasks and Criteria: A Mini-Development Activity* gives participants practice in designing tasks.

Activity 1.5—*Clapping Hands* is a compelling way to have participants generate conclusions about how tasks should be designed.

Activity 1.6—*A Comparison of Multiple Choice and Alternative Assessment* illustrates two task types (multiple choice and performance assessment) and asks participants to discuss their advantages and disadvantages.

Activity 2.3—*Ms. Toliver's Math Class* demonstrates tasks that are embedded in instruction.

Appendix A has many more examples of tasks.

Group Work. Some tasks are done individually, some are done in groups, and some require a combination of the two (see **Figure 3.6**).

Group work is usually incorporated into tasks if:

1. The student skill to be assessed is group collaboration. It's silly to design a separate task just to assess group collaboration. It is best done in the context of other tasks.
2. It's important to send a message that group collaboration is important.
3. Students need to engage in group work to "level the playing field." For example, let's say one wishes to assess ability to write. If the chosen task is one for which some students have more prior knowledge than others, not all students have an equal opportunity to shine. Group work allows for sharing background information that all students can then use to craft their essays. This technique was used in example L—History Paper.
4. Students need to interact in order to gather information needed for a subsequent individual task. For example, students might individually analyze survey data that is collected by the group.

Conclusion—Tasks. Tasks must be crafted to accomplish a plan. This plan outlines what you want to assess, the reason you are assessing it, the instructional message you want to send to teachers and students, the relationship you see between assessment and instruction, and your instructional philosophy. A summary of task design decisions is provided in **Figure 3.7.**

Figure 3.7
Summary of Task Design Decisions

- Open-Response (one correct answer) or Open-Ended (many possible correct responses)?
- Hands-On or Paper and Pencil?
- Written, Oral or Product response?
- Short Answer, On-Demand Performance Assessment, Project, or Portfolio?
- Group Work?

Ensuring High Quality Tasks—or—Reliability and Validity Made Easy

Overall Quality Considerations for Tasks. The central question in evaluating tasks is the extent to which the task elicits evidence of the intended skills or abilities. Did you ever give students an activity that didn't work out as expected? For example, you wanted students to write an Aesop's fable on how camels came to have humps, but instead they wrote informational pieces on the functions of humps. The point is a skill can't be assessed if the students don't use it. For example:

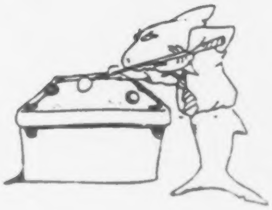
- If the goal is to assess problem solving, will the task on understanding the water cycle result instead in problem solving or regurgitation of memorized knowledge?
- Does the graphing problem give so many hints that students don't have to problem solve?
- Is the task on determining the amount of speeding near the school so novel that students really *will* have to employ experimental skills rather than test-taking skills or memorized algorithms?

In short, does one get what one wants?

Since most of the skills assessed using an alternative assessment are fairly complex (such as critical thinking, problem solving, communication, and group collaboration), tasks have to be "meaty." And since this type of task is usually complex and time-consuming, they should be worth the time and effort of students and teachers—meaningful, worthwhile, interesting, and examples of the best current thinking about educational theory. This set of overall quality considerations for tasks is summarized below:

1. Will the task(s) given to the student likely elicit the desired performances? Do you get what you think you're getting?
2. What is the extent to which the task recreates an "authentic" context for performance? Is the task of importance to the students? Do they find it interesting? Will it elicit "real" vs. "artificial" performance? Will assessment tasks engage and motivate students to perform to the best of their ability?
3. Are the tasks consistent with current educational theory and best practice? Is the task instructionally sound? Does it fit well with instructional goals and practices?
4. Have tasks actually been reviewed by content experts to judge content quality and authenticity?
5. Have tasks been pilot tested to shake out the bugs?

Sampling, Representativeness, and Generalizability Considerations for Tasks. Let's say that one wants to make a decision on how well a friend fishes. So he or she goes out with her on a beautiful sunny day about 3:00 p.m. to a lake of her choice. The friend catches no fish; the conclusion is that she is no good as a fisherman. Is this judgment valid?



Of course, we're hoping you'll say, "No!" because that 3 p.m. experience was only one sample of a very complex skill. To really make a determination of fishing ability one would need to sample his or her friend's fishing skills under different weather conditions, in different types of water, with various types of fish and equipment, and at different times of day.

Now, an educational example. Let's say that the goal of an assessment is to make judgments about individual student's ability to problem solve in mathematics. So, teachers develop ten tasks that they believe require problem solving and randomly distribute them in classrooms, with each student doing only one of the tasks. Can the results of this assessment be used to draw conclusions about the problem solving ability of *individual* students? Once again, we hope you're shouting, "No, of course not." Results on any single problem might be the result of a fluke; for example, the student might just have happened to be given a very similar problem last week in mathematics class.

Could the teachers, however, draw conclusions about the *group* of students as a *whole*? (That is, could they average performance together across all the students to see how they are doing as a group?) Maybe. It's unlikely that every single performance of *all* students is due to a fluke. Any fluke that results in a higher score is likely to be balanced out by a fluke that results in a lower score. The real issue in this case is not so much errors in assessing any individual student but whether the *sample* of ten problems adequately covers what the teachers mean by "problem-solving skills" and whether the math content of the problems adequately covers the content students should know. The question to ask to determine "adequate sampling" is: "If students are given five different, but similar, problems, would conclusions about their problem-solving ability be the same?"

This scenario illustrates the issues of sampling, representativeness, and generalizability. Is there a good enough range of tasks to justifiably draw a conclusion? Would student performance on other possible problems be similar to the performance on the particular set of tasks on the current assessment—in other words, could the teachers *generalize* from the current set of tasks to the whole domain of problem solving?

This is not a trivial issue. It is hard to predict from a student's performance on one task how he or she will do on a second task. For example, a student might develop a good design for one science experiment, but have less success with another.

Of course, if all that is wanted is to make a statement about how well students do *on the tasks on the current test*, issues of generalizability are not important. Just say, "On this test students did thus and so." Generalizability becomes an issue only when the goal is to generalize a conclusion *based on* a test. For example, one might want to say, "Based on this test, I judge that Alex is a competent problem solver." In this case generalizability is of paramount importance. The issue of generalizability is especially important when the conclusions one draws about students result in high-stakes decisions—for example, "Sevu has not shown competence in problem solving, therefore, he will not receive a high school diploma." In this case, we want to make very sure that the sample of problems upon which this decision is made represents the whole range (domain) of problem-solving skills and that there are enough problems to rule out flukes.

In the final analysis, issues of generalizability can only be answered by pilot testing—trying out the assessment and actually seeing how well performance generalizes to similar tasks and to other measures of the same skill. However, one should always be thinking about this issue when selecting or designing alternative assessments. **Always ask, "What's my best estimate of whether this set of tasks represents the domain well?" and "What's my best guess as to whether performance on this set of tasks allows me to draw the conclusions I want to draw?"**

The following questions summarize these quality considerations for sampling and the representativeness and generalizability of tasks:

1. How well do the tasks cover the skill area (domain)? Do they cover what is really important in the area of interest? What evidence exists to show that performance on these tasks is generalizable to other similar tasks and representative of the desired behavior?

2. Are enough samples of performance or displays of knowledge/skills/ behaviors collected to say with confidence that results adequately show what a student can do.

Bias and Distortion Considerations for Tasks. Bias and distortion occurs when there is something in the assessment that hinders the ability of students to demonstrate what they know and can do with respect to the skill or knowledge being assessed. Some of these things affect only individual students while others affect whole groups of students. In this section we'll deal with things that affect individual students. Features that can affect entire groups of students relate to serious equity issues and will be dealt with separately later under the general heading of "fairness."

Here are examples of features of performance tasks that might distort the picture we get of individual student achievement:

- What if a performance assessment requires role-playing, even though the ability to play a role has nothing to do with the ability or knowledge being assessed? In this case, the necessity of playing a role might inhibit the ability of a student to demonstrate the real skill being assessed, critical thinking, for instance.

- There might be a great deal of reading necessary to set up an assessment task, so much so that a student who doesn't read well might not understand the directions and, therefore, would be unable to do the task.

- An even more subtle type of bias distortion can occur when the performance requires *knowledge* extraneous to what is being assessed. For example, suppose students are asked to design a dog-run with a specified square footage. The goal is to demonstrate knowledge of the relationship between perimeter and area, yet in order to design a reasonable run, students need to know that an acceptable run has to be wide enough for a dog (i.e., six inches wide is not adequate) and large enough to provide exercise. Some students may not know this additional information. So, they might design dog runs that are technically correct, but not practical. Do they get scored down for that?

Related *Toolkit98* Activities and Chapters

Activity 1.5—*Clapping Hands* provides a hands-on experience during which participants generate for themselves many of the same quality guidelines as presented in this section.

Activities 3.1—*Performance Tasks, Keys to Success* and **3.6**—*How to Critique an Assessment* provide experience with friendly critiques of performance tasks using the guidelines outlined in this section.

Activities 3.7—*Chickens and Pigs*, **3.8**—*Questions About Culture and Assessment*, and **3.9**—*Tagalog Math Problem* illustrate bias and distortion issues that can stem from student backgrounds and cultures.

The rating form at the end of this chapter summarizes these quality issues.

Chapter 1 also discusses quality issues with respect to all assessments—alternative or not. Remember the "bubbles" on "sampling" and "bias and distortion?"

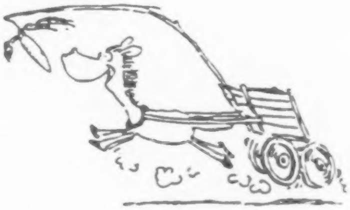
The following question summarizes quality considerations for tasks relating to bias and distortion:

> Are there things—either irrelevant or not critical to the skill actually being assessed—that might get in the way of students being able to demonstrate what they know and can do?"

Examples of such extraneous factors might include:

- Ability to read, write, or role play
- Ability to understand the context in which activities or tasks do or do not occur
- Personality (shyness, fear of public speaking, etc.)
- Extraneous background information required to do the task
- Physical limitations or conditions that interfere with performance
- Determining individual student ability through a group task

Performance Criteria–Design Options and Quality Considerations



Performance Criteria Design Options

We've looked at the first half of the performance assessment equation—tasks. Now, for the second half—performance criteria. Performance criteria are the methods given for evaluating the quality of student response to the task given them to do. Performance criteria are also called scoring guides or rubrics. There are several ways that performance criteria can be designed. This section outlines what we've found.

Task Specific vs. Generalized Criteria. Probably the major distinction between criteria types is the degree to which they are task-specific. **Task-specific** criteria means that the scoring guide can only be used for one, specific task because the rater looks for certain specific features in order to award points. For example:

> Scoring for **D** (*Solid Waste*)
>
> Sorting: 7 points for all correct. Subtract 1 point for each incorrect response. Sequencing: 1 point for ordering items correctly; 0 points for incorrect order. (See **Sample A.16** in **Appendix A**—*Sampler*.)

Or, task-specific scoring can be based on features of responses, as in the example below:

Scoring for *How Many Buttons?*

Students work together to estimate the number of buttons on people's clothing in the school. A preliminary activity had students estimate the number of buttons in their classroom. In a *high quality response* the student takes into "consideration the following aspects of the problem: (a) the data collected from the day before; (b) some estimate of the number of children per classroom; (c) the number of classes in the school; (d) variation of buttons among children of various ages; and (e) adults' buttons. Reasonable justifications are given for how each numerical value was chosen. Numerical values are put together with appropriate arithmetic processes. Steps are clearly detailed." (See **Sample A.14** in **Appendix A**—*Sampler*.)

Neither of these scoring guides could be used for a different task. For example, one couldn't use the scoring guide for the first example to score the student responses in the second.

Generalized criteria, on the other hand, can be used across similar tasks. Consider the following:

Scoring for *Mathematics Problem Solving*

Performance is scored using rubrics having four dimensions: conceptual understanding, procedural knowledge, problem solving/strategies, and communication. The same scoring guide is used for all problems given to students in grades 3, 5, 8, and 11. (See **Sample A.18** in **Appendix A**—*Sampler*.)

Scoring for *Mechanized Vehicle*

Performance is scored for thinking skills (conceptual understanding, creativity, and critical thinking), group collaboration, and communication skills (speaking and writing). These same criteria could be used for any group project. (See **Sample A.44**—*High School Physics*.)

This difference in the design of performance criteria relates to the degree to which one infers competence *indirectly* from performance on a task versus judging it *directly*. "Who cares?" you might be thinking. Although the importance of this distinction may not be entirely clear at first, it really does have far-reaching implications for how one views and uses alternative assessment. As an illustration, let's assume that one is designing an alternative assessment to measure mathematics problem solving.

Approach 1: Indirect inference. First, one would develop a math problem for students to solve. Then one would decide what a strong or weak performance would look like and assign points for these features. For example, getting a particular answer might get 3 points, having a particular graph with particular labels might warrant a point for each correct label, using particular descriptive phrases might warrant a point for each phrase used, and so forth with particular pictures, using particular measuring instruments in particular ways, etc. Here, problem-solving ability is not judged directly, but inferred from the number of points the student receives. For example, a high score on **Sample G**—*Aquarium Problem* is taken as evidence of problem-solving ability.

This is essentially the same logic as in multiple-choice tests.

Approach 2: Direct judgment. Contrast this to **direct judgment** of the skill being assessed, in this example, problem solving. In this approach one would design tasks to elicit problem solving, but then, instead of scoring the response right/wrong or assigning points for correctness or specific features, one would directly judge the quality of the problem solving using **generalized criteria**. For example, to get a high score, a response might have to be clear and focused, have used a good strategy, be insightful, be complete, include relevant diagrams or examples when appropriate, focus on all important parts of the problem, etc. The criteria don't require any *specific* strategy but, instead, sound thinking regardless of the strategy used. Likewise, the criteria don't require any *specific* diagram, only that if a diagram is used, it helps and is correct. In this case, the *same* criteria are applied to *all* problems solved by students

C Related *Toolkit98* Activities and Chapters

See **Activities 2.1**—*Sorting Student Work*, **2.2**—*How Can We Know They're Learning?* and **2.5**—*How Knowledge of Performance Criteria Affects Performance* for demonstrations of the power of generalized criteria to sharpen our notion of the outcomes we have for students and provide a vocabulary for discussing features of quality performance that can be used in the classroom for instruction and student self-assessment.

and problem solving is judged directly by the rater.

This difference is important because it relates to our vision of how alternative assessment fits into the educational enterprise. For example, the indirect approach (emphasizing task-specific scoring) is more reliable and faster for large-scale assessment because raters don't have to analyze each student's work for evidence of problem solving; all they have to do is look for specific diagrams, calculation or explanations, and give points for their presence. All the judgment work has already been done—someone else has already decided that on problem #A, good problem solvers would have a particular diagram, calculation, or explanations.

The more direct approach (emphasizing generalized scoring procedures) has important advantages for instructional use in the classroom because it *forces assessors to do the analysis themselves*. Making judgments about the qualities that make a performance successful helps assessors internalize the criteria for success. This becomes especially powerful if the assessors are teachers and students.

The following sentiments explain the link to instruction.

> *Scales, criteria and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new materials even when they do not have the criteria in front of them. [This procedure is] two times more effective than [other] techniques.*
> (Hillocks, 1986)

> *Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses.* (Lucas, 1992)

📖 Reference Box

G. Hillocks, *Research on Written Compositions*, 1986, Urbana, IL: National Council of Teachers of English.

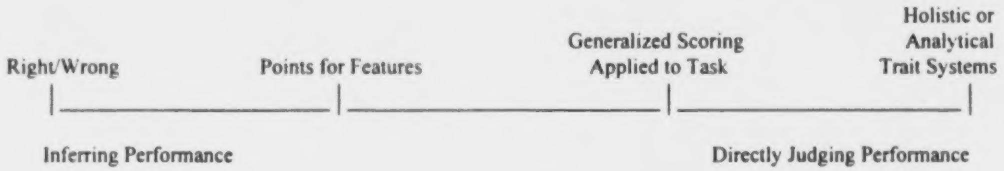
C. Lucas, *Introduction: Writing Portfolios—Changes and Challenges*, in K. Yancey (Ed.), *Portfolios in the Writing Classroom*, 1992 Reston, VA: National Council of Teachers of Mathematics.

Nidhi Khattri, Alison Reeve, and Rebecca Adamson, *Studies of Education Reform, Assessment of Student Performance*, U.S. Department of Education, 1997.

In other words, making direct judgments of the skill being assessed helps teachers and students conceptualize "what good looks like" across situations. When scoring is instead done for specific features of tasks, raters only get to know what "good" looks like on a single problem; they don't have to generalize, transfer or apply knowledge to a new situation. Therefore, direct rating of performance using generalized criteria is better as an instructional tool. In fact a new national study makes this very point:

The influencing of instructional practices to date has been served most powerfully by generic rubrics. (OERI, 1997, p. XX)

Of course, life is never black and white. What has been described are two ends of a continuum:

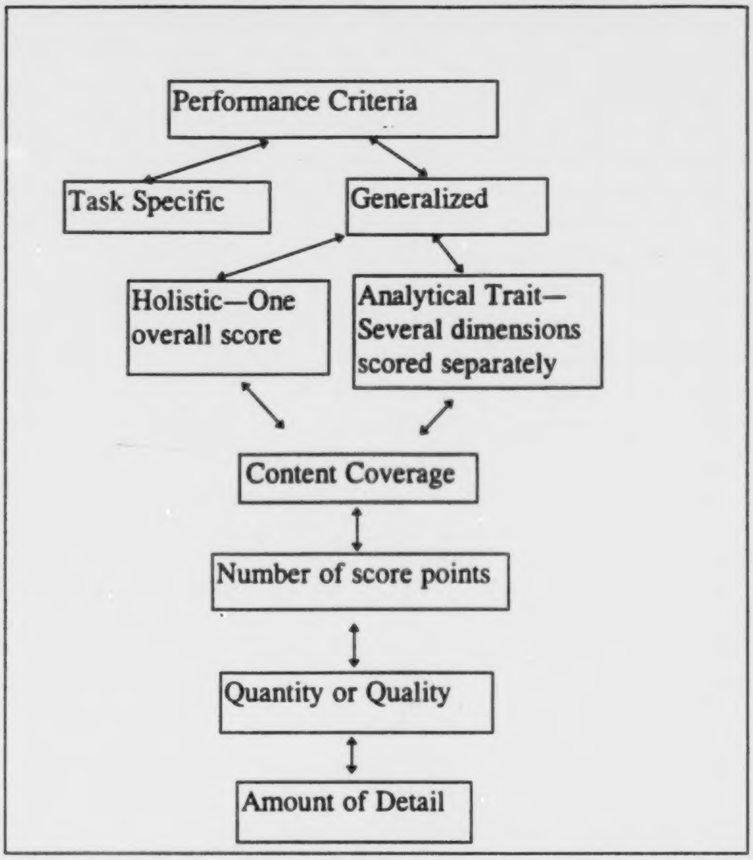

Right/Wrong
Points for Features
Generalized Scoring Applied to Task
Holistic or Analytical Trait Systems
Inferring Performance
Directly Judging Performance

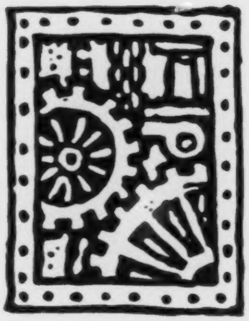
In actuality, alternative assessment approaches fall all along this continuum. One "happy medium" can occur in cases where a general set of scoring guidelines is developed and then systematically applied to each task to develop specific scoring guides based on points. This approach can actually help teachers and students generalize so that eventually they aren't dependent on specific scoring guides. Such an approach is used by several states.

To be successful as an instructional tool, however, the link between the specific example (e.g., "has a picture showing a whole cut into four equal parts") and the general criterion (e.g., "uses appropriate graphic representations") has to be clear, and the ultimate goal has to be understanding the general criteria and applying them to new situations.

Now, let's just focus on the general criteria for awhile; forget all about task specific criteria. Within the broader category of generalized criteria, there are several other decisions to be made: holistic or analytical trait, number of score points, quantity or quality descriptions of score points, and the amount of detail to use. These are depicted in **Figure 3.8**. (Some of these might also apply to task specific criteria but, in this chapter, we're more concerned with general criteria because of their superior instructional value.)

Figure 3.8
Decisions About Performance Criteria


Performance Criteria
Task Specific
Generalized
Holistic—One overall score
Analytical Trait—Several dimensions scored separately
Content Coverage
Number of score points
Quantity or Quality
Amount of Detail



Holistic vs. Analytical Trait Generalized Scoring. **Holistic** scoring entails giving one overall judgment about the quality of the response as a whole. **Analytical trait** scoring entails giving separate scores for various important dimensions of performance. (This would be similar to assessing how well each of the parts in the machine to the right works. Each of the parts is essential to the whole, but the gears might work better than the whistle.) A good example of analytical trait criteria is the Six-Trait Model (**Sample A.10** in **Appendix A—*Sampler***), which assesses six essential aspects of successful writing—ideas, organization, voice, word choice, sentence fluency, and conventions. Each dimension, or trait, would get a separate score, showing a profile of strengths and weaknesses. By contrast, a holistic rubric for writing would have the rater lump all of these together to come up with a single, overall score.

The advantage of holistic scoring is that it is quicker. Therefore, it is used in large-scale assessment when the main goal is to get a relatively quick overview of student performance. The advantage of analytical trait scoring is that it is more diagnostic. Therefore, it is used when the goal is either to provide diagnostic information to teachers for planning instruction or when the goal is to use criteria as a way for students to understand the nature of quality and to self-assess.

Content Coverage. What do the criteria define as important features of a strong performance? What are indicators of quality? What should teachers *not* take into account when judging quality? These features make up the content of the criteria.

Content is important for two reasons. First, since performance criteria provide the basis on which student responses will be judged, they should cover *all* the most important dimensions of performance. What sense would it make, for example, to judge the quality of a student's problem solving skills only by whether he or she arrived at the right answer? Judgments of the quality of a student response might also need to include the sophistication of the approach taken, the efficiency of the procedures chosen, evidence of checking to make sure the answer is correct, etc. Ultimately, students need to come up with reasonable answers, but they also need to do it efficiently.

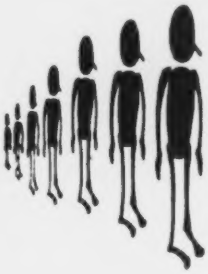
The second reason content of criteria is important is that teachers and students, understandably, concentrate their efforts on what is included. What students and teachers see is what will be taught and learned. Therefore, it's important that criteria deal with all the valued dimensions of performance, not just those that are easy to define and score. For example, consider communication in mathematics. What instructional messages are sent to teachers and students if criteria leave out the student's ability to clearly communicate what he or she did? If communication isn't assessed, teachers might not emphasize it in instruction and students won't know that it's important to learn.

Assessment in writing offers another example. Conventions in writing are essential. If one can't decipher a piece of writing, one can't judge it's quality in other ways. But, conversely, conventions aren't everything. Most teachers also value strong ideas, organization that supports the point being make, voice that sets the appropriate tone and style for the writing, words that precisely convey what is meant, and sentences that are not only correct but add to the flavor of the piece. What might happen to writing performance if any of these were left out of the criteria for judging writing quality?

Just as it's essential to include the more important dimensions of performance, it's equally important to leave out the trivial. A two-minute speech might be desirable, but what happens if a student gives an oral presentation that is fantastic, but only 90 second long?

The content of the performance criteria provides the final definition of local content standards. What gets assessed is the de facto definition, so we all better be really sure that the criteria adequately cover all that is important and *only* that which is important.

Nature of the Scoring Scale—Number of Score Points. Once one has decided on the major approach to scoring (task-specific or generalized) and decided to have one score for a student performance or several, one needs to decide how points will be awarded. Will there be a four-point scale (or 2, 3, 5, 6, 7, 8, or whatever)? And then, what will constitute a score of 4, 3, 2, or 1?



First, the number of score points. In the scoring schemes we've seen, rubrics have from 4 to eight levels. More than that makes it hard for raters to distinguish between levels. ("Remind me again what's the difference between a 57 and a 58.") Fewer than four levels and there is little leeway to show progress. Some people like an odd number of points (so that there is a middle balance point of strengths and weaknesses), and others favor an even number of points to force an overall judgment of "on the better side" or "on the worse side." Personally we favor five- or six-point scales as a balance between the need to have many points to show progress and not *so* many points that teachers have difficulty distinguishing differences in quality.

Nature of the Scoring Scale—Quantity or Quality. Now, on to deciding the features that will distinguish levels. Believe it or not, there *are* variations in the philosophy behind how to define each level of the trait(s) to be scored. Here are two:

Scoring Based on Quantity

In some assessments, a score depends on the number of some feature that is present. For example, one assessment in the **Appendix A—** *Sampler* has four traits (**Sample A.41**—*Writing in Chemistry*)— concepts, argumentation, use of references to the text, and misconceptions. Three of these are based on "counts." That is, you count up the number of misconceptions, concepts used, or proportion of text detail used, and the score depends on the number you find.

Scoring Based on Quality

In other assessments, judgments are based on the quality of the responses. For example, one feature of a high performance might be that "the student successfully proposes an explanation which clearly shows a relationship between data and final conclusions." The corresponding feature of a weak performance is that "the final conclusion is unrelated to the tasks performed." It doesn't matter how many conclusions the student comes to as long as the conclusions drawn are good ones.

In general, rating scales based on quantity should be avoided because they very easily can become arbitrary and might send the wrong messages about what is important. For example, one way to distinguish between levels in performance is to count the number of errors, or number of resources used, or some other attribute of the performance. This could lead to unintended negative consequences such as students using easier words to avoid spelling mistakes, or loading a report with references, whether they are most relevant or not. After all, one good conclusion should be worth more than 10 poor conclusions.

Amount of Descriptive Detail. Another way criteria can differ is with respect to the amount of detail provided to help define each level.

Scoring with Sketchy Descriptions of Terms

In *Solid Waste* (**Sample. A.16**), to get a high score, the student must give a "strong, workable solution." The rater—teacher or student—receives no assistance in understanding what it means to have a strong, workable solution. What does a strong, workable solution look like?

Scoring with Some Descriptions of Terms

In **Sample A.18**—*Math Problem Solving, Oregon*, to get a high score on "Procedural Knowledge," the student "uses principles efficiently while justifying the solution, uses appropriate mathematical terms and strategies, solves and verifies the problem, and uses mathematical principles and language precisely." While still not completely definitive, this provides the rater with greater assistance in knowing what to look for in the response.

Scoring with More Complete Descriptions of Terms

The Six-Trait Model (**Sample A.10**) was developed to be used with students and, so, attempts to carefully define both what is meant by each trait scored and what performance looks like at various levels of quality. For example, the trait of "voice" is defined as: "the writer coming through the words, the sense that a real person is speaking to us and cares about the message. It is the heart and soul of the writing, the magic, the wit, the feeling, the life and breath. When the writer is engaged personally with the topic, he/she imparts a personal tone and flavor to the piece that is unmistakably his/hers alone. And it is that individual something—different from the mark of all other writers—that we call voice."

Strong voice occurs "when the writer speaks directly to the reader in a way that is individualistic, expressive and engaging…the tone and voice give flavor to the message and seem appropriate for the purpose and audience,…narrative writing seems honest, appealing, and written from the heart,…expository or persuasive writing reflects a strong commitment to the topic, and an effort to bring the topic to life by anticipating the reader's questions and showing why the reader should care or want to know more."

Weak voice occurs when "the writer seems indifferent, uninvolved or distanced from the topic and/or the audience. As a result, the writing is lifeless or mechanical,…the writer speaks in a kind of monotone that flattens all potential highs or lows of the message,…the writing may communicate on a functional level, but it does not move or involve the reader,…the writer does not seem sufficiently at home with the topic to take risks,…."

We like detail!!! First, it's easier for raters to be consistent when description increases. This is important for large-scale assessment. Even more important, however, is the communication potential of descriptive criteria. If teachers tell students, for example, that their solutions have to be "workable" will that help them know what to do? Wouldn't it be more helpful to let them know that "workable" means a balance between being practical (in terms of time, knowledge, and money) and being effective (in terms of really solving the problem); that effectiveness has to be demonstrated by more than just opinion (they need data on the effectiveness of other similar solutions); and that the technology needs to be currently available to implement the solution. Can you think of additional details that would be useful for students?

The point is that good quality performance criteria can be more than just a tool for giving a score on a test if they are designed wisely. They can help teachers define expectations for students and develop a shared vision of quality. This requires detail—definitions, indicators, and examples of what we mean. Be wary when someone says, "We have to make all this fit on one page." The goal is not to make it fit on one page but to adequately describe what we mean. This might take ¼ page, but it might instead take 6.

Conclusions on Performance Criteria. There are many types of performance criteria under development. The kind of scoring mechanism chosen should directly relate to what is to be assessed, one's purpose for assessment, and how one sees alternative assessment fitting into the educational enterprise. Holistic and task-specific scoring are often used for large-scale assessment because they are faster. Analytical trait performance criteria with lots of description may be the best choice to develop a common vision of the goals for students and to maximize the chances that alternative assessments will be useful instructional tools in the classroom.

⊂ Related *Toolkit98* Chapters and Activities

Activity 1.9—*Going to School* lets workshop participants try out different kinds of performance criteria and discuss their advantages and disadvantages. It also demonstrates the necessity of having performance criteria for large-scale assessment.

Activity 2.1—*Sorting Student Work*, and **Activity 2.2**—*How Can We Know They're Learning?* illustrate how to develop performance criteria and show how they can be tools for defining our goals for students. In **Activity 2.4**—*Is Less More?* Participants develop criteria for a task they write. **Activity 2.5**—*How Knowledge of Performance Criteria Affect Performance* illustrates the importance of criteria in helping to define the task.

Activity 3.3—*Performance Criteria—Keys to Success* talks about the essential features of performance criteria if they are to be used as tools for communication in the classroom.

Appendix A has many more examples of performance criteria.

Figure 3.9 summarizes design considerations for performance criteria:

Figure 3.9
Design Considerations for Performance Criteria

• Task specific	*or*	• General
• Holistic	*or*	• Analytical Trait
• Quantity description of score points	*or*	• Quality description of score points
• Number of score points		
• Amount of detail in description of score points		

Ensuring High Quality Performance Criteria—or—Reliability and Validity Made Easy Once Again

Here is a summary of quality questions to ask about criteria:

1. Do the criteria cover what is important? Are the concepts, processes, skills, etc. that you wish to measure appropriately addressed by the criteria? For example, if you want to assess problem solving, are all the important dimensions of a good problem solution covered?

2. Are the criteria stated clearly enough so that raters can agree on meanings? Is the amount of detail just right? Are there examples of student work to illustrate the criteria?

Additional questions to ask if the performance criteria are intended to be useful instructional tools are:

3. Are criteria general? If we are measuring "big" outcomes, can criteria be used to assess those skills across tasks?

4. Are the criteria analytical trait?

Quality—The Rest of the Story on Reliability and Validity

There are several important considerations that apply to alternative assessments in general or that apply to both tasks and criteria. We'll finish up this chapter with a brief discussion of these.

Content/Skill Coverage and Proper Match to Assessment Method

Content/Skill Coverage. Assessment coverage should closely match the goals of the curriculum, and the assessment device should cover the entire range of that content. The student knowledge and skills covered by an assessment should be important ones, and content should represent enduring themes or significant knowledge, not trivialities. In five years, what would you want your current students to remember or be able to do based on your time with them?

It's easy to say that content coverage must be appropriate, but discovering what an assessment actually covers is not as simple as it might first appear. Sometimes, for example, the developer makes a written statement about coverage (e.g., "this assessment measures critical thinking, attitude toward mathematics, and problem solving"), but it is impossible to match these statements to what the students are asked to do on the assessment or what the performance criteria cover. Or, there is no statement about coverage and one has to figure out what the assessment covers from the performance tasks and criteria. *Ideally, statements about coverage match what the tasks ask students to do and what is covered by the criteria. Therefore, when we ask about what the assessment assesses, we need to consider tasks and criteria in addition to any written statements of coverage made by the developer.*

➲ Related *Toolkit98* Chapters and Activities

Activity 1.5—*Clapping Hands* provides a hands-on experience during which participants discern for themselves characteristics of high quality performance criteria.

Activity 3.3—*Performance Criteria—Keys to Success* provides criteria for high quality criteria, complete with models and anchors.

Activity 3.6—*How to Critique a Assessment* provides practice critiquing performance criteria using the four questions in this section.

394

Alternative Assessment is the Best Match. Not all the learning targets for students are best measured using alternative assessment as amply discussed in **Chapter 1**. Many people believe that the most productive use of alternative assessment is to address "big" learning targets such as (1) problem solving, reasoning, critical thinking and attitudes; (2) complex applications of knowledge and skills to new settings or situations; or (3) the processes that students use to complete tasks. To assess such complex learning outcomes, many indicators are needed, including performance assessments.

Related *Toolkit98* Chapters and Activities

Chapter 1 extensively discusses appropriate learning targets for students and matching targets to assessment methods.

Activity 1.2—*Clear Targets, What Types are These?* asks participants to consider whether a good mix of learning targets is covered.

Activity 1.6—*A Comparison of Multiple-Choice and Alternative Assessment* discusses when different types of assessments should be used.

Activity 1.7—*Target-Method Match* asks participants to match targets to assessment types.

Activity 1.8—*Sam's Story* illustrates the need for multiple-measures of student skills.

Activity 1.10—*Clear Targets and Appropriate Methods, View From the Classroom* asks participants to self-assess how well they can match targets to methods.

The following quality questions summarize important quality questions about content alignment and importance:

1. Is it clear what goals, skills and/or content are to be covered on the assessment?
2. Are these goals, objectives or targets best assessed using an alternative assessment? Why or why not?
3. Does the assessment address and embody valued student outcomes and avoid irrelevant or unimportant content and skills?
4. Does the assessment deal with enduring themes or significant knowledge within or across disciplines?
5. Do statements of coverage, task content, and performance criteria match up?

Fairness and Rater Bias

Some of the most commonly asked questions in workshops on assessment are: "How do we handle ESL, special education or Title I students? Do we hold them to the same standard? Do we give them the same tests? What accommodations and modifications are OK?"

All of these questions relate to issues of fairness. As teachers we want to be as fair to all students as much as possible. Teachers want to ensure that every group of students has an equal opportunity to show what they know and can do, and that language and cultural differences will not give us a distorted picture of a student's capabilities.

In the section on performance tasks we talked about possible sources of bias and distortion that affect individual students. In this section we're talking about sources of bias and distortion that can affect entire groups of students. The topic of group bias is so important we wanted to give it special attention.

Potential sources of group bias come in many sizes and shapes and are often very subtle. We all know that students who don't speak English very well won't do very well on any test in English regardless of what it's trying to measure. But, how many know, for example, that in some cultures children do not discuss their opinions (as in American Samoa)? If a group discussion is used to assess, say, critical thinking skills, a biased view of the achievement of students from this cultural group might result.

An even more subtle potential form of bias is the context in which a skill is assessed. An obvious example is a math task set in the context of a football game. The context becomes a group biasing factor if particular groups of students know less about football than other groups of students and this might affect their ability to do the math problem. If boys, in general, have more experience with the game and more knowledge of its structure and rules than do girls, then the task could be biased in favor of boys. What might happen to students from another culture who know nothing about football? Even if no knowledge of football is necessary for understanding the problem itself, students unfamiliar with the context are put at a disadvantage by having to deal with multiple tasks—the task intended by the assessor and the task of making sense of a problem in an unfamiliar area when students are keenly aware that they are missing pieces of knowledge that may prove vital.

There are also less obvious examples. Tasks viewed as "content free" or "everyday settings" can reflect hidden assumptions about daily life and experiences that differ vastly across socioeconomic, ethnic, and gender groups. For example, people eat green bananas in American Samoa, so the answer to the question, "What color are bananas?" is green, not yellow. Folks in rural settings know that if there are five birds on a limb and you scare one away, the rest will go too. So the answer as to how many are left is "0" not "4." Similarly, urban students might have only seen chicken in packages at the store, so questions on the number of legs on 10 chickens and a pig might not elicit knowledge of multiplication.

Group bias may exist in the rater as well as in the task itself. For raters, attitudes, beliefs, and values, often held unconsciously, may be reflected in their judgments. For example, a rater may let handwriting affect a judgment of content, or preconceptions about abilities of boys and girls may influence scoring decisions. These biases are often unconscious. Most raters will not even be aware that they have biases.

Social/cultural norms and experiences, social class, gender, and ethnicity all influence the knowledge and resources that students bring to assessment tasks. Recognizing biases in assessment tasks and developing ways to overcome such biases is a difficult but critical part of designing and evaluating assessments, whether alternative or traditional.

The following quality questions summarize what should be asked about an alternative assessment to ensure fairness:

1. Do the content or context of the task or assumption about expected performances reflect knowledge, experiences, and values that are equally familiar, acceptable, and appropriate for all groups of students?
2. Does the assessment tap knowledge and skills all students have had adequate opportunity to acquire?
3. Is the assessment as free as possible of cultural, ethnic, or gender stereotypes?
4. Do students understand the language of the assessment?

Related *Toolkit98* Chapters and Activities

Chapter 1 discusses potential sources of bias and distortion in assessment results.

Fairness and equity issues are explored in **Activities 3.7**—*Chickens and Pigs,* **3.8**— *Questions About Culture and Assessment*. and **3.9**—*Tagalog Math Problem.*

These issues also frequently arise during **Activities 1.5**—*Clapping Hands* and **3.6**—*How to Critique an Assessment.*

Important rater bias questions:

1. Is some feature of the performance (e.g., handwriting, poor spelling, pet peeves) influencing how another, supposedly independent, feature is judged?
2. Does knowledge of the type of student performing the task (e.g., gender, ethnic heritage, curriculum track) influence judgments?

3. Does knowledge of individual students affect judgment of performance? (e.g., "Susie always does such good work, I'll give her the benefit of the doubt. Heather always does such poor work, someone else must have done this for her.")

Consequences

Any assessment may have unanticipated side effects. Boy, don't we all know that. A positive side effect might be that a student's attitude toward a subject area improves because the assessment was fun to do. We're all more accustomed, however, to seeing the negative side effects—restricting curricula to what can be easily assessed; communicating unintended messages about power, control, and social roles/status in the classroom; or providing narrow and restrictive images of the nature of knowledge in a particular subject area.

The history of testing is replete with examples of good intentions gone awry: tests that were supposed to improve the educational system but that actually resulted in narrowing instruction to basic skills; or multiple-choice tests, originally developed to provide more "objective" measures of achievement that actually resulted in perpetuating the idea that there is always a right answer to any problem. Good intentions are clearly not enough. Cognizant of past problems, assessments must be designed that minimize unintended negative consequences.

Science and mathematics education have been particularly plagued by several myths that have been reinforced by traditional assessments. These include:

- One right answer exists to all mathematics and science problems.
- Computation is central to mathematics and science.
- "Knowing" is more important than being able to do (for example, being able to pick a correct answer is more important than being able to carry out a lab activity).
- Science and mathematics are primarily done with pencil and paper.
- Being able to recognize technical terms is more important than understanding concepts.

📖 Reference Box

Nancy Kober, 1993. *What We Know About Science Teaching and Learning.* National Education Knowledge Industry Association—NEKIA, (formerly Council for Educational Development and Research), (202) 429-5101.

- Some external authority is the only valid judge of whether an idea in science or mathematics is acceptable or "true." (Nancy Kober, 1993, p. 59)

These beliefs have a powerful influence on student learning, use of knowledge, interest in and attitude toward subject areas, and access to knowledge in particular areas. Assessment that communicates such notions can have a strong negative impact on student lives. This does not mean that factual knowledge is unimportant. Emphasizing reasoning without concepts and facts is impossible.
Clearly, a balance is needed.

When examining the consequences of an assessment, consider the following quality questions:

1. What will this assessment communicate to students, teachers, and others about:
   - What is important to teach? Will there be effects on instruction and curriculum?
   - What is important to learn? For example, a student might say, "I don't have to worry about problem solving because all the problems on the test come from homework."
   - What is the nature of the discipline(s) being assessed and the nature of knowledge in those subject areas? For example, is an assessment sending a message that mathematics is primarily computation?
   - Who has the authority to judge what is acceptable knowledge and performance in these areas?
   - What images of learning, and of life, are being communicated?
   - How should we teach? Does the assessment exemplify good instruction?
2. What effects might the assessment have upon the lives, attitudes, emotions, self-image, career options, etc. of the teachers, students, and parents involved?
3. Is the assessment worth the instructional time devoted to it? Will students learn something from the assessment itself?

⊂ Related *Toolkit98* Chapters and Activities

A discussion of how assessment communicates what is valued can be found in **Chapter 1.**

Discussion of unintended consequences of assessment often arise as part of **Activity 1.5**—*Clapping Hands* (effects on students), **Activity 1.9**—*Going to School* (effects on instruction), **Activity 3.6**—*How to Critique an Assessment*, and **Activity 2.1**—*Sorting Student Work (*effects on instruction).

4. Does the assessment provide information that is relevant to the decisions that will be based on it?

5. Will the assessment be perceived by students and teachers as a valid indicator of student competence in the particular area being assessed?

Cost and Efficiency

An assessment should be practical. The consequences of the assessment should be worth the cost and effort required to do it. By necessity, assessment design always requires a series of trade-offs. For example, observing students doing laboratory activities may provide a more accurate picture of student lab skills than examining laboratory notebooks. However, examining notebooks is more efficient. Does the increase in accuracy justify the increase in time needed for direct observation? The issue is whether the loss in validity is acceptable when compared to the increase in efficiency.

Impacts and effects can include more than just obtaining good data. Involving students in rich, engaging tasks and in self-assessment can promote the very learning that is being assessed. Also, involving teachers in the design and scoring of alternative assessments can be an extremely powerful professional development experience. This makes the additional cost of alternative assessment more worthwhile.

In evaluating assessments, it is important to identify the relevant trade-offs and determine the acceptability of gains or losses in validity, instructional usefulness, reliability, and efficiency. Although addition of authentic, alternative, and performance-based tasks usually raises the cost of an assessment, an assessment that meets the criteria described in this chapter will more likely be worth the investment. Questions that must be considered are:

1. Is the assessment cost-efficient? Are the effort and the resources required to do the assessment worth the results and consequences?

2. Is the assessment "practical" or "do-able"? (For example, does the assessment require so much time, equipment, space, etc. as to preclude use? Are there safety or privacy concerns?)

Chapter Summary and Conclusion

There is considerable activity in the development of alternative assessments in all subject areas, at all grade levels, and at all levels of education (classrooms to the federal government). The samples we have collected show a variety of approaches. In all these examples, the same reasons are given for changes in assessment: the need to better reflect current theories of instruction, changes in instructional goals for students, etc. However, the link between the general rhetoric and how the rhetoric plays out in the specifics of individual assessment design is often not clear. For example, if everyone is saying we need to better model good instructional practice, why do some assessments only use manipulatives while others only use traditional paper and pencil word problems? Or, why, when people are assessing math problem solving, do some groups use task-specific scoring while others are developing general criteria? We need to step back and reconsider all our design features to make sure they really *do* match what we are trying to accomplish.

Pay special attention to the evidence of validity and reliability put forth by the developers of an alternative assessment. Many times there is none! An assessment is always more trustworthy if the developers have attended to the quality issues outlined in this chapter—whether the tasks elicit the correct student performance, how consistently responses can be scored, whether tasks or scoring are biased toward any particular group, whether the results can be generalized to other similar tasks, and the impact of the assessment on students, teachers, curriculum, and instruction. In general, except for well-funded large-scale assessment projects, very little technical information is available on the alternative assessments we have seen. In some cases, pilot testing is under way. In other cases, little or no attention seems to have been paid to this important topic.

Alternative assessment can be a powerful tool for learning if we do it correctly. We hope the considerations and activities in this chapter will help teachers be wise consumers and good developers of alternative assessments.

The attached form summarizes the quality considerations discussed in this chapter.

Figure 3.10
Alternative Assessment Evaluation Form

	Yes	Somewhat	No
1. Content/Skill Coverage and Correct Method	3	2	1

The assessment:

- Clearly states skills and content to be covered
- Correctly uses alternative assessment to measure these skills and content
- Avoids irrelevant and/or unimportant content
- Deals with enduring themes or significant knowledge
- Matches statements of coverage to task content and performance criteria

2. Performance Criteria:	3	2	1

- Include everything of importance and omit irrelevant features of performance
- State criteria clearly and provide samples of student work to illustrate them
- Are stated generally, especially if the intent is use as an instructional tool
- Are analytical trait, especially if the intent is use as an instructional tool

3. Performance Tasks:	3	2	1

General:

- Elicit the desired performances or work;
- Recreate an "authentic" context for performance
- Exemplify good instruction
- Are reviewed by others (students, peers, experts)

Sampling/Representativeness/Generalizability:

- Cover the content or skill area well; results can be generalized
- Sample performance in a way that is representative of what a student can do

Bias and Distortion:

- Avoid factors that might get in the way of student ability to demonstrate what they know and can do

	Yes	Somewhat	No
4. Fairness and Rater Bias	3	2	1

Performance Tasks:

- Have content and context that are equally familiar, acceptable, and appropriate for students in all group
- Tap knowledge and skills all students have had adequate time to acquire in class
- Are as free as possible of cultural, ethnic, or gender stereotypes
- Are as free as possible of language barriers

Performance Criteria and Rater Training:

- Ensure that irrelevant features of performance do not influence how other, supposedly independent, features are judged
- Ensure that knowledge of the type of student does not influence judgments about performance quality
- Ensure that knowledge of individual students does not affect judgments about performance quality

5. Consequences	3	2	1

The assessment:

- Communicates appropriate messages
- Results in acceptable effects on students, teachers, and others
- Is worth the instructional time devoted to it; students learn something from doing the assessment and/or using the performance criteria
- Provides information relevant to the decisions being made
- Is perceived by students and teachers as valid

6. Cost and Efficiency	3	2	1

The assessment:

- Is cost efficient—the results are worth the investment
- Is practical/"do-able"

Activity 3.1

Performance Tasks—Keys to Success

Purposes:

1. To learn the dimensions along which performance tasks differ
2. To develop skill in identifying the characteristics of quality performance tasks
3. To practice matching assessment tasks to the purpose for the assessment, one's instructional philosophy, and one's vision of integrating assessment and instruction

Uses:

This is an advanced level activity that can be used in **Chapter 3** to illustrate different task types, how task type relates to purpose and learning target to be assessed, the advantages and disadvantages of different types of tasks, and what constitutes a good quality task. Prerequisites might include (a) an activity on the rationale for alternative assessment (e.g., **Activities 1.1**—*Changing Assessment Practices...*, **1.6**—*Comparing Multiple-Choice and Alternative Assessment*, or **1.12**—*Assessment Principles*); (b) **Chapter 1** text or **Activity 1.7**—*Target Method Match*; (c) **Chapter 2** text or an activity, such as **2.3**—*Ms. Toliver*, illustrating integrating assessment tasks into instruction; and (d) one of the activities providing a gentle introduction to quality issues, such as **1.5**—*Clapping Hands* or **1.11**—*Assessment Principles.*

Rationale:

A variety of performance task types are being developed—everything from short answers that look like multiple-choice with the choices taken off to complex, long term, group projects. These differ in complexity, the degree to which they require integration of skills or test individual skills in isolation, amount of collaboration, length of time required, presentation and response modes, amount of student choice, and the number of correct answers. It is necessary for educators to consider which of these options will accomplish their assessment purposes.

BEST COPY AVAILABLE

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A3.1,O1**—*Performance Tasks Keys Purposes*, **A3.1,O2**—*Sample Oral Presentation Task*, and **A3.1,O3**—*How Tasks Differ*
- **Handouts A3.1,H1**—*Criteria For Performance Tasks* and **A3.1,H2**—*Mathematical Applications Test*
- Sample performance tasks that illustrate various dimensions of task design. Samples and guidance on their use are included in the **Activity 3.1**—*Index* in **Appendix A**—*Sampler*. Run these off as handouts as well as prepare them as overheads.

Time Required:

60-90 minutes

Facilitator's Notes:

1. (10 minutes) As needed, review the purposes for the activity (**A3.1,O1**—*Purposes)* and what performance tasks are.

   There are two parts to a performance assessment—tasks and criteria.

   We use the term "tasks" to refer to whatever activity, project, prompt, problem, or assignment is given to the students to do. The quality of the student response to the task is what is assessed using the criteria (scoring guide, rubric).

   As needed, brainstorm examples of tasks. Examples of tasks: oral reading, writing a research paper, doing a science lab, group discussion, doing a worksheet on fractions, taking a multiple-choice test. Some assessment tasks are developed specifically to assess student skills; other tasks are those that occur as part on normal on-going instruction. The point is that assessment doesn't have to be conceived of as something that occurs at the end of, and is separate from, instruction.

   There are two important quality considerations when designing performance tasks—will it elicit the right performance and will there be anything about the task that might get in the way of any group of students showing what they can do. The first question is covered below in steps 2-5. Steps 6-7 cover the second question.

2. (5 minutes) *The main purpose of the task in an alternative assessment is to elicit the right performance while avoiding sources of bias and distortion—things that can cause us to make an incorrect judgment of student skill level. By "eliciting the right*

performance" we mean that we get from students what we wanted to get. Here are some examples of times that the task did not elicit the right performance. (Come up with your own examples, or use the examples below.)

a. On a third grade writing assessment, the students were given the prompt: "You've heard the story about how the camel got its hump. Think of something in nature and write a story about how it got that way." The intent was to assess imaginative writing by asking students to write an "Aesop's Fable." What was actually obtained were thousands of student essays on why camels have humps. Imaginative writing could not be assessed because the students wrote informational pieces instead.

b. Would a math problem assess problem solving if students had just completed several problems just like it? In this case, the problem might be eliciting recall of information, not problem solving.

c. Two sixth-grade teachers asked students to choose a novel and participate weekly in a book group. They planned to assess student reading comprehension, in part, by means of a final project—an oral presentation. What they got was a good assessment of oral presentation skills, but not a good assessment of reading comprehension skills.

The point here is that if you want to assess critical thinking, for example, you have to have a task that results in students thinking critically. Or, if you want to assess, for example, group process skills, you have to give students a task that enables them to work cooperatively.

Ask the participants to provide examples from their own classrooms of tasks or activities that did not work as intended.

3. (10 minutes) *Whatever the nature of the student learning target being assessed, whether simple or complex, the task needs to elicit the right performance or it can't be assessed. For simple learning targets (e.g., can students spell 15 words), the task can be simple. For complex learning targets, such as critical thinking, problem solving, writing, etc., you need tasks that are complex. We need "meaty" tasks or the right performance will not be elicited.*

   To get this point across, ask participants to think of a successful activity they have used in their classroom with students, and what made it so. Give participants 3-4 minutes to come up with a list of attributes that made their classroom activity successful. Then convene the large group and make a list on an overhead. Common responses at all grade levels are: student choice, active, clear instructions, engaging, relevant, organized, students know the purpose, enough time, several ways to approach the task, proper resources, and accommodates different learning styles.

   The interesting thing about the list that teachers generate is that this is the same list used by many authors to illustrate what is meant by "authentic" assessment. For example, see **Handout A3.1,H1**—*Criteria for Performance Tasks*. Point this out to participants.

Good assessment tasks that will elicit complex, meaty performances from students look a lot like good meaty instructional tasks.

4. (10 minutes) Use **A3.1,H1**, the "meatiness" scale, to evaluate and discuss several performance tasks. Take each example and ask participants two questions: *What skills are elicited by this task? How meaty is the task on the meatiness scale?* Keep repeating the mantra *simple skill, simple task; complex skill, complex task.*

   Use examples you have discussed previously; or, if this is a stand-alone activity, a thought-provoking set of tasks are listed in the **Activity 3.1**—*Index* in **Appendix A**—*Sampler*. This set was selected to represent a range of "meatiness." Therefore, these task types can be though of as "anchors" for the "meatiness" scale. **Handout A3.1,H2**—*Mathematical Applications Test* can be used to demonstrate how a multiple-choice test measures up on the "meatiness" scale. **Overhead A3.1,O2**—*Sample Oral Presentation Task* can be used to illustrate the weak end of the meatiness scale.

5. (15 minutes) Go back over the task examples to see what features tend to make them more or less "meaty." Teachers need to become expert at designing performance tasks that will elicit the right performances on the part of students. So, they need to think about what features lend themselves to meatiness. For example, heavily scaffolded tasks do not assess problem solving ability. **Handout A3.1,O3**—*How Tasks Differ* can be used to focus the discussion.

   On **Handout A3.1,O3**, you might highlight the features that *tend to make* a task "meaty." (Remember that one always has to take into account what one is trying to assess when making the call about "meatiness." For example, if you want to see whether students can write an essay, you must have them write an essay; therefore, choice on ways to respond might not be relevant.)

   - One right answer vs. MORE THAN ONE
   - Written, oral and/or visual instructions/activities: MAXIMIZE THE VARIETY OF WAYS TO DO THE TASK (as much as possible)
   - Amount of choice on how to respond—written, oral and/or visual: MAXIMIZE THE VARIETY OF WAYS TO REPOND (as much as possible)
   - Format—short answer, essay, project, portfolio: IN GENERAL MORE COMPLEX TASKS (essay, project, portfolio) ARE MEATIER
   - Specifying each skill that the student needs to use or ALLOWING STUDENTS TO CHOOSE WHATEVER SKILLS THEY THINK APPROPRIATE TO SUCCESSFULLY COMPLETE THE TASK
   - Individual and/or group work: EITHER

- Timed or UNTIMED
- STUDENT CHOICE ON WHICH TASK TO COMPLETE

6. (10 minutes) The second characteristic of a good task is to avoid pitfalls—sources of bias and distortion. Some of these probably came up during the discussion of tasks above. Participants may have said things like:

   a. "This requires a lot of writing. What about students who don't write well?"

   b. "This requires a lot of reading. What about students who are ESL?"

   c. "Will students who have had experience with carnivals do better than students who haven't?"

   d. "How will group work affect the judgment of ability of individuals?"

   When these comments arise, note that this is what is meant by "sources of bias and distortion" (and other terms like reliability and validity) and keep a running list. Then, at this point, have participants brainstorm additions to the list. You could even categorize the list into major sources of bias and distortion: equity, fairness, and bias; extraneous interference; sampling.

7. (10 minutes) Go back to the tasks reviewed previously for "meatiness" and critique them for "potential sources of bias and distortion."

8. (30 minutes) An extension would be to have participants attempt to develop a performance task that would elicit a valued student skill and avoid sources of bias and distortion.

Performance Tasks—Keys to Success

Purposes:

1. To learn the dimensions along which performance tasks differ
2. To develop skill in identifying characteristics of quality performance tasks
3. To practice matching assessment tasks to the purpose of the assessment, one's instructional philosophy, and one's vision of integrating assessment and instruction

Overhead A3.1,O1

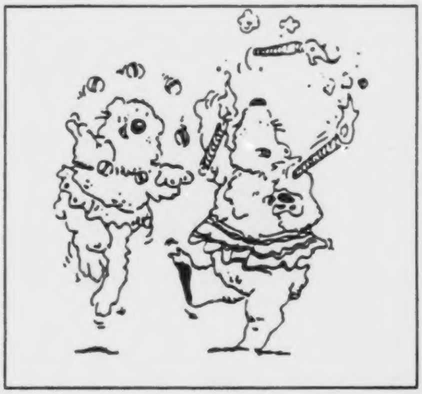
Sample Oral Presentation Task

You have 5 minutes to prepare a 3 minute persuasive talk. You may choose one of the topics below or develop your own topic.

Gun control
Right to life
School uniforms
Free trade zones
Environmentalism
Urban growth districts
Doctor assisted suicide
Raising the voting age to 21

How Tasks Differ

- One right answer vs. more than one
- Written, oral and/or visual instructions/activities



- Amount of choice on how to respond—written, oral and/or visual
- Format: short answer, essay, project, portfolio
- Specifying each skill that the student needs to use or allowing students to choose whatever skills they think appropriate to successfully complete the task
- Individual and/or group work
- Timed or untimed
- Student choice on which task to complete
- What else?

411

Overhead A3.1,O3

CRITERIA FOR PERFORMANCE TASKS
"Meatiness Scale"

Here is a list, adapted from one developed by Steve Leinwand and Grant Wiggins, of some important characteristics that we should look for in tasks. We will find it helpful to return often to this list, reviewing the quality of the curriculum and assessment activities we present to students.

Very "Meaty" ←——————————→ Not "Meaty"
5 4 3 2 1

ESSENTIAL	• The task fits into the core of the curriculum. • It represents a "big idea."	vs.	TANGENTIAL
AUTHENTIC	• The task uses processes appropriate to the discipline. • Students value the outcome of the task.	vs.	CONTRIVED
RICH	• The task leads to other problems. • It raises other questions. • It has many possibilities.	vs.	SUPERFICIAL
ENGAGING	• The task is thought-provoking. • It fosters persistence.	vs.	UNINTERESTING
ACTIVE	• The student is the worker and decision maker. • Students interact with other students. • Students are constructing meaning and deepening understanding	vs.	PASSIVE
FEASIBLE	• The task can be done within school and homework time. • It is developmentally appropriate for students. • It is safe.	vs.	INFEASIBLE
EQUITABLE	• The task develops thinking in a variety of styles. • It contributes to positive attitudes.	vs.	INEQUITABLE
OPEN	• The task has more than one right answer. • It has multiple avenues of approach, making it accessible to all students.	vs.	CLOSED

Adapted from "Criteria for Performance Tasks," from *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*, edited by Jean Kerr Stenmark. Reston, VA: National Council of Teachers of Mathematics, 1991. Reprinted with permission.

Handout A3.1,H1

412

Mathematical Applications Test

1. John and Sue are going to the movies. If they spend $12.00 for tickets, $3.50 for popcorn and $2.50 for drinks, how much money will they have left from a $20.00 bill?

   a. $12.00

   b. $8.00

   c. $6.00

   d. $2.00

2. An advertisement states, "Of doctors surveyed, 90 percent rate no other brand of sleeping aid better than Baye's." Which of the following could you conclude from this statement?

   a. Baye's sleeping pills may or may not be better than other brands.

   b. Baye's sleeping pills are the best.

   c. Almost all the doctors in the country prefer Baye's sleeping pills.

   d. Baye's is probably worse than other brands because advertisements always lie.

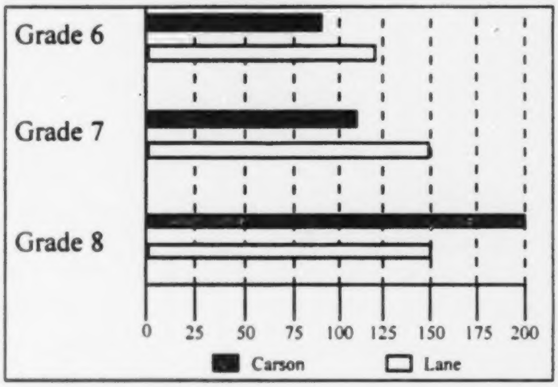
3. Quan Soo and Eddie have 9 baseball cards and 15 basketball cards. Eddie bought 6 more baseball cards. Which of the following statements is true?

   a. There are more baseball cards than basketball cards.

   b. There are more basketball cards than baseball cards.

   c. There is an equal number of basketball and baseball cards.

4. How many more eighth graders are there at Carson School than at Lane as shown on the graph below?



   a. 50 students

   b. 100 students

   c. 150 students

   d. 200 students

Handout A3.1,H2

413

Activity 3.2

Spectrum of Assessment Activity (Revisiting Bloom's Taxonomy and Alternative Assessment)

Purposes:

1. To find the strongest matches between assessment methods and kinds of thinking (using Bloom's Taxonomy)
2. To practice the "can-do" verbs that indicate levels of Bloom's Taxonomy
3. To explore assessments that elicit complex thinking skills

Uses:

This is an intermediate level activity that can be used in **Chapter 1** to illustrate the need to match assessment method to the learning targets being assessed, or in **Chapter 3** to help teachers design the tasks that assess all levels and types of student thinking skills. Prerequisites might include (a) an activity on the rationale for alternative assessment (e.g., **Activities 1.1**—*Changing Assessment Practices...*, **1.6**—*Comparing Multiple-Choice and Alternative Assessment*, or **1.12**—*Assessment Principles*); (b) **Chapter 1** text or **Activity 1.7**—*Target Method Match*; and (c) **Chapter 2** text or an activity, such as **2.3**—*Ms. Toliver*, illustrating integrating assessmenty tasks into instruction.

Rationale:

This activity will make participants aware that different types of assessments elicit different types of thinking from students. Multiple-choice and short answer questions tend to require simpler thinking on the part of students than performance assessments, for example. Participants will be given the opportunity to review Bloom's Taxonomy in a non-threatening manner in order to review the various types of thinking skills we want to develop in students.

414

Next, they will look at a series of assessment tasks (questions and problems) and rank them along a spectrum according to their cognitive complexity. The goal is to point out that it is usually the alternative assessment that requires higher level thinking. Participants can then review the questions they themselves ask students and rate them for cognitive complexity.

As with other activities in the *Toolkit98*, we want participants to understand the match between targets (student learning goals) and assessment methods. Simple targets require simple tasks while complex targets, like higher-order thinking skills, require complex tasks, like those involved in a performance assessment.

This activity is broken into two parts: Part A is "Revisiting Bloom's Taxonomy"; Part B is "The Spectrum of Alternative Assessment."

PART A–Revisiting Bloom's Taxonomy

(This review is based on the old version of Bloom's and not the proposed revision.)

Materials:

- Overhead projector, screen, Post-it® notes, markers, chart paper
- **Overheads A3.2,O1**—*Spectrum of Assessment Activity Purposes,* **A3.2,O2**—*Bloom's Taxonomy Pyramid,* **A3.2,O3**—*Discussion Questions (Parts 1 and 2),* **A3.2,O4**—*Application Discussion Questions*
- **Handouts A3.2,H1**—*Choral Reading,* **A3.2,H2**—*Bloom's Taxonomy Blocks*, and **A3.2,H3**—*Using the Question (Assessment Task) Quilt With Students*
- Other manipulatives: flower blooms (cut from paper as signs for each level of Bloom's Taxonomy that can be worn around the neck), thumb tacks, masking tape, wolf hat, bow ties (felt or paper), 1-3 different versions of the "Three Little Pigs," and play sunglasses (large frames)

- Prepare a "Question (Assessment Task) Quilt" on chart paper that looks like the following:

Question (Assessment Task) Quilt for the "Can Do" Verbs					
Recall					
Application					
Analysis					
Synthesis					
Evaluation					

- Prepare cards containing each verb on **Handout A3.2,H2**—*Bloom's Taxonomy Blocks*. Make the verbs for each level of Bloom's Taxonomy a different color. For example, all the "Analysis" verbs might be green and the "Evaluation" verbs might be blue.

- Hang up five sheets of chart paper with a horizontal line dividing each into two equal parts. Bloom's Taxonomy levels of Recall, Application, Analysis, Synthesis, and Evaluation should be written across the top of the sheets. (See the diagram below. Recall includes both Knowledge and Comprehension.) Pull the bottom of the chart paper up toward the top and tape it to cover up the words until you are ready for them.

Recall	Application	Analysis	Synthesis	Evaluation

Time:

90 minutes to 2 hours

Facilitator's Notes:

1. (5 minutes) Explain the purposes for this activity. Use **A3.2,O1**—*Spectrum of Assessment Activity Purposes*.

2. (5-10 minutes) Introduce the Three Little Pigs Activity.

   a. You might say: *"Participants are here today to study higher-order thinking skills. We decided to use some difficult books to work with and hope you will not be turned off." Hold up the three different versions of "The Three Little Pigs" and ask if anyone remembers the basic story?*

   b. Ask each participant to write (on a Post-it® note) a question one might ask a child (aged 8-14) about the story of "The Three Little Pigs." (No talking or sharing for now).

   c. Unveil the chart paper with the Taxonomy Levels and say, *"Place your question on the chart paper above the line in the level of Bloom's Taxonomy to which the question belongs."* (Many participants will feel very uncomfortable with this, but reassure them that mistakes will not be pointed out).

3. (5-10 minutes). Ask participants if there are any particular areas where there were more questions posted? If so, why? Ask participants if they were uncertain of the level of question they asked and whether they remember Bloom's Taxonomy? Some information to point out: In 1991, three authors (Risner, Micholson, and Myhan) determined the percent of questions, based on Bloom's Taxonomy, that were used in popular fifth-grade science textbooks: 58% were Knowledge Level, 36% were Comprehension Level, (combined as Recall—94%), 6%—Application, all other levels—0%). The rest of the activity is intended to refresh participants' memory of Bloom's Taxonomy to ensure higher level questions are asked in the classroom.

> **Reference Box**
>
> Gregory P. Risner, and others, 1991. ***Levels of Questioning in Current Elementary Textbooks: What the Future Holds.*** ERIC ED# 344 770

4. (45-60 minutes) Review Bloom's Taxonomy

   a. Ask for seven volunteers to play the parts in the choral reading, using **Handout A3.2, H1**—*Choral Reading,* and distribute their props:

      - One volunteer represents the WOLF (prop: Wolf hat)
      - One volunteer represents Lenny (prop: sunglasses)
      - The other five represent the Blooms (prop: paper flower blooms that have each level of Bloom's Taxonomy spelled out indicating which level of Bloom's they represent, i.e., recall...)
      - The rest of the group represents the Chorus (prop: felt bows that can be used as a bow-tie or hair-ribbon)

b. Read the choral reading, "The Big, Bad, Wolf Meets the Blooms." (**A3.2, H1**—*Choral Reading*). After the story is read, review each level of the Taxonomy with the participants by asking them to explain the meaning of each level by relating the levels to the choral reading context. Use **Overhead A3.2,O2**—*Bloom's Taxonomy* as a guide. (Also explain to the participants that this choral reading models an alternative instructional approach they can use with their students.)

c. Hand out a card containing a verb from **Handout A3.2,H2**—*Bloom's Taxonomy Blocks* to each person and a new posting note. Have each participant write a new question or assessment task about "The Three Little Pigs," using the verb on the card. Ask the participants to place their new question or task on the chart paper below the horizontal line in the level of Bloom's to which the question belongs.

d. Post the "Question (Assessment Task) Quilt." Have participants place their verb on the Quilt in the appropriate level. (In the "Materials" section of the facilitator's notes, you were directed to color code the verbs so that they'll produce a pattern when placed on the Quilt. For example, all the verbs in the "Analysis" row might be green.) Participants will catch on to the coding scheme and some might then want to move the question or task they wrote using the verb to another level on the Bloom's Taxonomy sheets. The point here is the importance of using the verbs as a way to ensure that teachers daily ask questions in the classroom at higher levels of Bloom's Taxonomy.

(Note: There might be a discussion of factors other than the verb that determines the levels of Bloom's Taxonomy addressed by a question or task. For example, multi-step tasks makes a question more complex, as does asking, "Why?")

e. Ask participants to discuss the questions on **A3.2,O3**—*Discussion Questions.*

f. Using **Handout A3.2,H2**—*Bloom's Taxonomy Blocks* with the "can do" verbs and possible products, review the alternative assessment possibilities with the participants. (Make sure that participants recognize that when writing questions or tasks for students, the verb used is the action the student does and not just a word in the question or statement. For example, if a student is asked to contrast something, you need to specify how this contrasting will be completed— in graphic form, in chart form, in outline form, in a summary, or what? What will the product be?)

- Make sure that participants realize that all levels of Bloom's Taxonomy must be used (Recall-Evaluation); each level builds on the other.
- Emphasize the "products vs. verbs" on the blocks in **A3.2,H2** and tell them to create more "can do" verbs and products, etc.

g. Ask for comments, questions or concerns about higher-order thinking skills.

5. Extensions:

a. (30 minutes) Apply Knowledge—Give each group of 6 (or so) teachers a column of verbs from the Questions (Assessment Task) Quilt. Have them construct a question or task using their column of verbs after doing one or more of the following.

- Watch a video segment (National Geographic is good)
- Bring in lesson notes (If told before the workshop)
- Use state benchmarks or objectives
- Use textbooks or other resource books

Ask participants to share their questions and other assessment tasks and discuss the points on **Overhead A3.2,O4**—*Application Discussion Questions.*

b. (15 minutes) Using the Question (Assessment Tasks) Quilt with students: Remind teachers that higher-order thinking questioning works best in classrooms when students are aware of Bloom's Taxonomy and they can sometimes help develop the questions. It is okay if students help. HOW? Pass out and review **Handout A3.2,H3**—*Using the Question (Assessment Task) Quilt With Students.*

🗀 Reference Box

Other taxonomies besides Bloom's can be used to generate assessment questions and tasks that measure the full range of thinking skills we want to develop in students. For example: Robert J. Wilson, 1996, *Assessing Students in Classrooms and Schools*, pp. 21-28. Allyn & Bacon Canada, 1870 Birchmount Rd., Scarborough, Ontario, M1P 2J7.

Jay McTighe, Maryland Assessment Consortium, c/o Urbana High School, 3471 Campus Dr., Ijamsville, MD 21754, (301) 874-6039.

Part B—Continuum (Spectrum) of Assessment

Materials:

- **Overheads A3.2,O5**—*Spectrum of Assessment*, **A3.2,O6**—*The Key to Student Success*, **A3.2,O7**—*Choosing the Right Tool*, **A3.2,O8**—*Range Indicator for Tasks*, **A3.2,O9**—*Instructions*, **A3.2,O10**—*Summary*, **A3.2,O11**—*Final Thoughts*
- **A3.2,H4**—*High School English Questions*, **A3.2,H5**—*Elementary/Middle School Science*, **A3.2,H6**—*High School Social Studies Questions*, **Handouts A3.2,H7**—*Elementary Social Studies*, **A3.2,H8**—*Elementary Social Studies-Blank*

Set Up:

Run the handout sheets (by subject) on different color paper and cut the questions apart (with letters attached). Place these cut out sets into envelopes.

Time:

90 minutes

Facilitator's Notes:

1. (5 minutes) Talk about a spectrum (like the electromagnetic spectrum) and how we can create a spectrum of assessments from "Simple" to "Complex" (roughly corresponding to "traditional" to "alternative") depending on our need. Relate this spectrum to the levels of Bloom's Taxonomy from Recall to Evaluation using **Overheads A3.2,O5**—*Spectrum of Assessment*, **A3.2,O6**—*The Key to Student Success*, and **A3.2,O7**—*Choosing the Right Tool*. You might say: "*When we move from traditional to alternative assessment, can we say we move from recall to a higher level on Bloom's taxonomy? We know that some multiple choice, true/false and matching can be used to assess various levels of Bloom's Taxonomy, but how often do we spend time making sure our "traditional" methods assess the higher-order thinking skills important for students?*"
2. (5 minutes) Ask participants to group themselves into pairs.
3. (5 minutes) Display overhead **A3.2,O8**—*Range Indicator for Tasks*, to explain how assessments in any content area can illustrate moving from "simple to complex" or "recall to evaluation." Ask: What are some things we need to look for when writing assessment questions and tasks to assess the higher end of Bloom's Taxonomy? A few things to think about are: (a) the 'can do' verbs from Bloom's used in the question that reflect what a student is asked to do, (b) the number of things a students is asked to do in a

question (multi-stepped tasks), and (c) allowing for many types of Gardner's Multiple Intelligences.

4. (5 minutes) Allow pairs of teachers to pick an envelope with elementary/high school social studies, elementary/middle school science, or high school English questions depending on their comfort with the subject. Place **Overhead A3.2,O9**—*Instructions* on the screen and go over the directions with the whole group. Directions are:

   - Place each task on a spectrum from the most "simple" (more traditional) assessment to the most "complex" (alternative) assessment—decide if the order goes from "recall" to "evaluation."
   - Each group should try to reach a consensus on their spectrum order but a difference in opinion is allowed.

5. (20 minutes) Allow sufficient time for participants to sort their tasks.

6. (20 minutes) Ask each group to report out their "spectrum" order. This can be done orally or on index cards. If time is a factor, ask groups to report their most traditional (or recall level) task and their most alternative task (or most evaluative level). Ask them to clarify their orders.

7. (10-20 minutes) Discuss ways to use this activity at the classroom level to help teachers design better assessments.

8. (5 minutes) Use **Overheads A3.2,O10**—*Summary* and **A3.2,O11**—*Final Thoughts* to summarize important points.

9. Wrap-up: Evaluate session (identify levels of Bloom's Taxonomy) WHAT WILL YOU TRY?

Spectrum of Assessment Activity

Purposes:

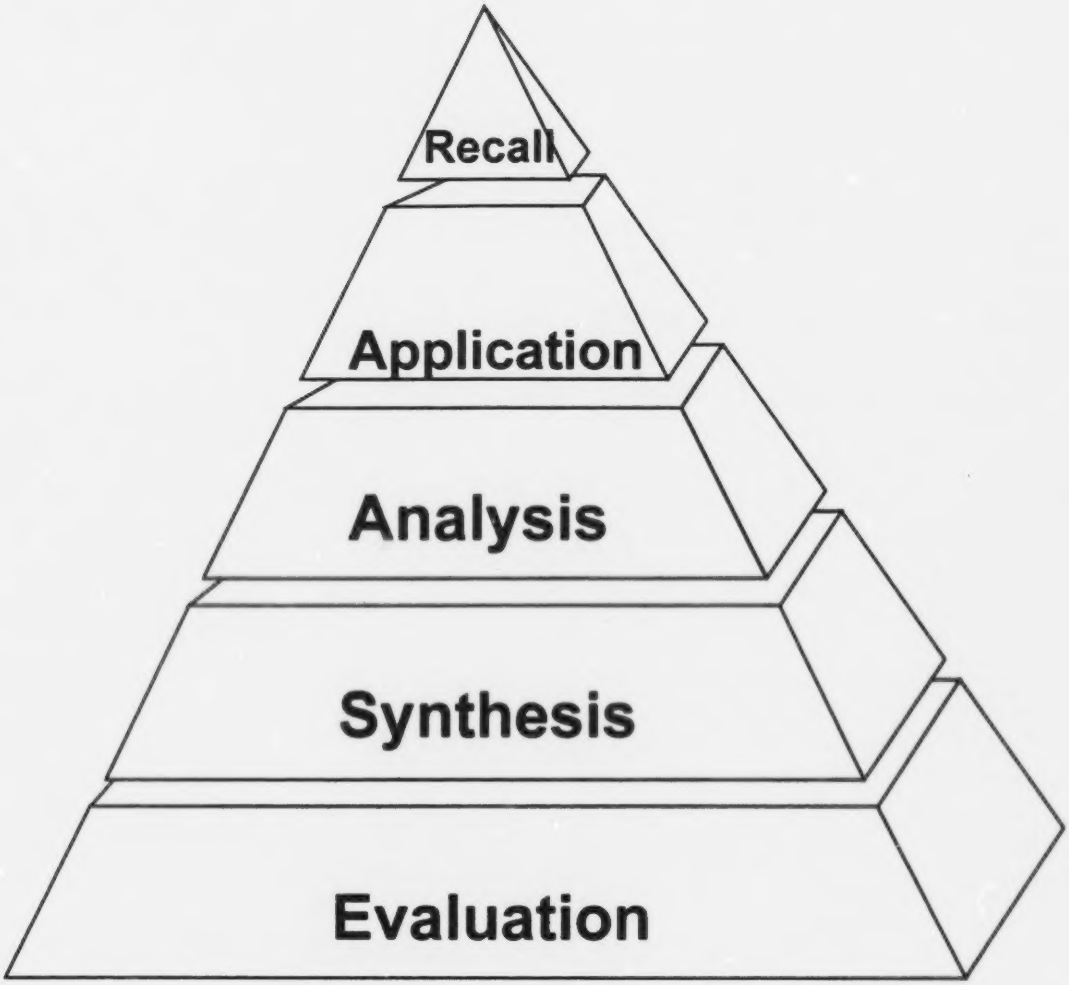
1. To find the strongest matches between assessment methods and kinds of thinking (using Bloom's Taxonomy)

2. To practice the "can do" verbs that indicate levels of Bloom's Taxonomy

3. To explore assessments that elicit complex thinking skills

Overhead A3.2,O1

Bloom's Taxonomy



Recall

Application

Analysis

Synthesis

Evaluation

Overhead A3.2,O2

Discussion Questions, Part 1

1. In what way do "can do" verbs help one develop targeted questions? Explain.

2. Did you have to think more to create a question or task when you had a specific verb to use? Explain.

3. How might these verbs be used in the classroom to develop questions and other tasks for assessments and lessons? Explain.

Discussion Questions, Part 2

4. How are these "can do" verbs reflected in local objectives, standards, or benchmarks? Explain.

5. How can you use the "can do" verbs to align local objectives to instruction and assessment? Explain.

6. When using higher-order questions and tasks with students, do you find traditional assessment methods limited? Explain. What types of alternative assessments could be used with specific levels of Bloom's Taxonomy? Explain.

Application Discussion Questions

1. Which questions or other assessment tasks would generate an alternative product?
2. Which questions and tasks are the most creative and would allow for a more creative response?
3. Which questions or tasks are the most subjective to score?
4. Which questions and tasks would students be more motivated to do?
5. Which questions or tasks would be the most difficult to grade on a test? Why?

Overhead A3.2,O4

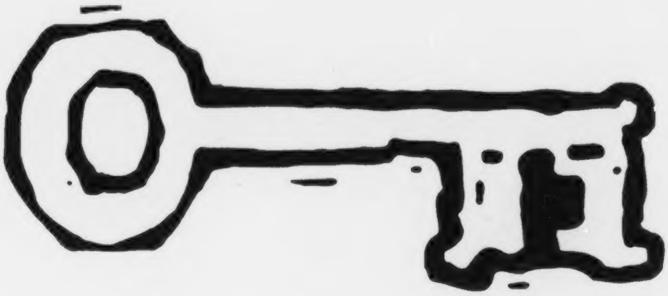
Spectrum of Assessment

A look at sample tasks..

Overhead A3.2,O5

The key to student success

If we alter assessment tasks in small ways, we can open them up and move them along the assessment spectrum to provide truer pictures of what students know and can do, and assess more complete skills.



Overhead A3.2,O6

Choosing the Right Tool

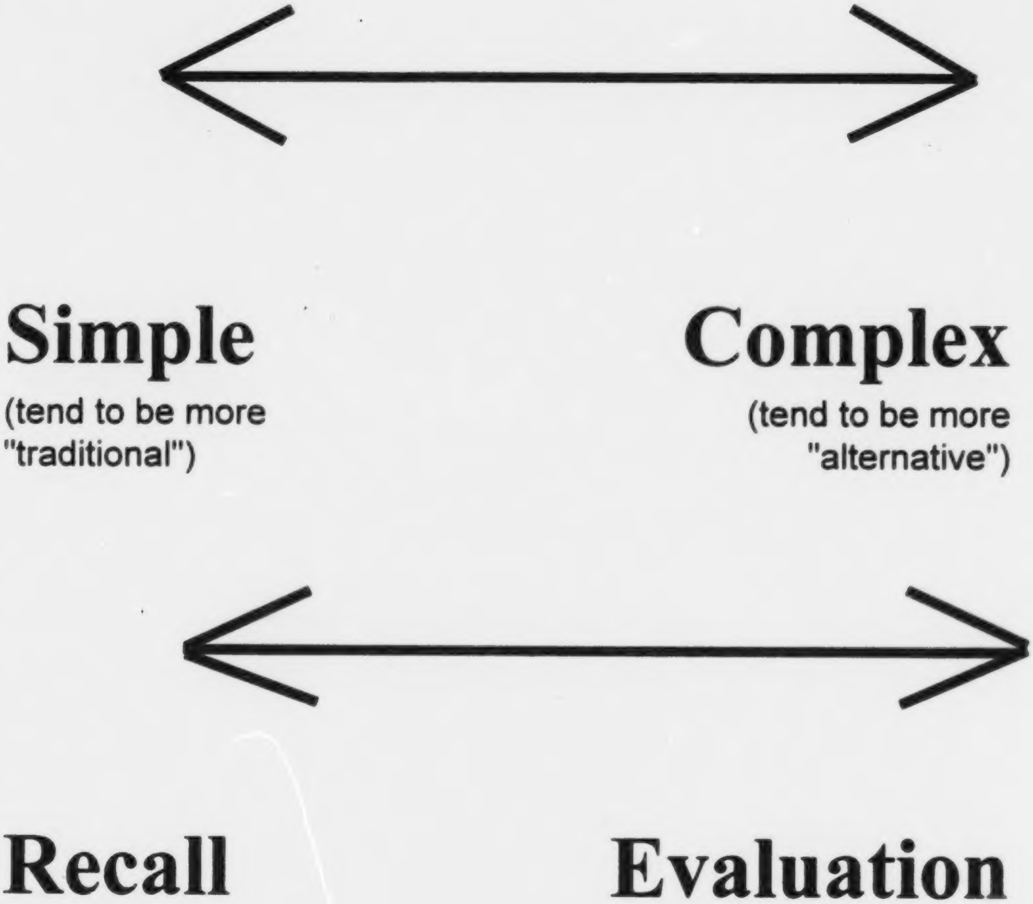
- Assessments are tools we use to gauge what students know and can do.
- It is very important to choose assessment methods that match what we want to assess so that we can make appropriate inferences about a student's performance.

Overhead A3.2,O7

429

Range indicator for tasks...

Assessment Spectrum



Simple
(tend to be more
"traditional")

Complex
(tend to be more
"alternative")

Recall

Evaluation

Overhead A3.2,O8

Instructions

1. **Reach a consensus on where you would place each task on the spectrum.**

   Complex (tasks in which):
   - students are more engaged with target knowledge and understanding
   - opportunities are present for students to show what they know and can do in a variety of ways

2. **Decide on characteristics that determine placement along the spectrum.**

3. **Note major areas of agreement or disagreement about placement.**

Overhead A3.2,O9

Summary, part 1

- Students need experience with complex oral and written questions,…so
- Teachers should use a variety of question types,…which is an example of
- Matching student learning targets to assessment methods.
- Alternative assessments match best to complex targets, like the higher end of Bloom's Taxonomy.
- Prior to testing, read questions over carefully to check for clarity.

Overhead A3.2,O10, p.1

Summary, part 2

- Questions requiring written responses are great, but go over expectations for quality answers with students prior to giving assignments (rubrics are great for open-ended types of questions).
- Prepare questions and other assessment tasks prior to class activities to emphasize various levels of student thinking.
- Students should practice writing good assessment questions.

Final Thoughts

- Small changes in the way a question or task is posed can provide a bigger picture of what students know and can do
- The best cameras with the best lenses can take poor pictures if used incorrectly. To get the best "snapshots" possible, choose your tools wisely to create the best pictures of what your students know and can do.

Overhead A3.2,O11

The Big, Bad Wolf Meets the Blooms

A CHORAL READING TO REVIEW BLOOM'S TAXONOMY

BY NANCY MCMUNN, ADAPTED FROM BARBARA HISE'S CHORAL READING

Characters: **LENNY:** (Wolfs psychologist-wearing glasses)
Wolf: (Wolf ears)
The Blooms: Recall, Application, Analysis, Synthesis, and Evaluation
The Chorus: Rest of the participants; everyone gets a part

The play opens with the Wolf seated center front, alone with his head in his hands. The chorus can be seated in a semi-circle fashion around the Wolf.

Wolf: Why, oh why is it that I never do anything right? I have tried for years to destroy the Pig family and I can never succeed.

Chorus: Poor Wolf. He always does everything wrong.

(Enter Lenny)

Lenny: Hello Wolf. Did you blow down the brick house yet and get rid of the Three Little Pigs?

Wolf: Naw, I got mad when the 2nd Little Pig fled from the stick house. I didn't catch him...again....He keeps running to the brick house and I haven't figured out a way to blow that house down.

Chorus: Poor Wolf. He can't do anything right.

Wolf: Hey, cut that out (facing the chorus). I'm not in the mood for teasing! I've got problems.

Lenny: Come on and join the **Bloom Society**, Wolf. They'll teach you to think like the 3rd Little Pig.

Chorus: Go on, Wolf. Listen to Lenny.

Wolf: What society? Besides, I can't. I never succeed at anything I try.

Lenny: Of course, you can succeed. You aren't the Scarecrow character in *The Wizard of Oz*. You do have a brain. You just don't know how to use it!

Handout A3.2,H1, p. 1

Wolf: Well, I might as well be, because I know I've never been taught to use my brain. I sure don't seem to be able to solve my own problems. **Life used to be so simple**—that is, *before the Three Little Pigs*.

Chorus: Don't know how to use the brain you've got!

Lenny: Well, I know how to fix that. I'm going to introduce you to my *brain fixers*, the **Bloom quintuplets**! They always help me with my problems....

Chorus: The **Bloom quintuplets**?

Lenny: Hey, **Blooms**! Come on in! I have a friend that needs your help.

(Enter: the five Blooms, each clearly labeled with a sign around his/her neck.)

(As Lenny introduces each skill, they step forward and take a bow.)

Lenny: Wolf, I want you to meet my secret skills, the thinking skills: **Recall, Application, Analysis, Synthesis, and Evaluation.**

Wolf: What **good** are they going to do me? I don't even know what their names mean and what kinds of skills they are?

Recall: Then I'm the first skill you need. I'm **Recall**. I'm easy to use. You won't have to do much thinking at all, in fact, you're ready to take a test now. What material did the 3rd Little Pig use to build his house?

Wolf: **Recall**, that's easy! He used bricks. I don't understand, so what good are you?

Lenny: **Recall** will not help you much without the next step.

Application: And that's me. My name is **Application**. When you take something you have learned and put it to use, you are applying knowledge.

Wolf: Let's see. I learned that the Three Little Pigs built their houses out of straw, sticks, and bricks and that was my first level of thinking. I only used **Recall** of my knowledge. Now, I know your name is **Application**—but what can you do for me?

Application: I bet the 3rd Little Pig used **Application** when he built his house out of the pile of bricks.

Lenny: Well, yes, because if you think about it all the Little Pigs used **Application** when they constructed their houses. Or, you used **Application** when you demonstrated that you could blow down the straw house because you recalled that the house was constructed from straw and straw was easy to blow away.

Wolf: Oh, I think **Application** is smarter than **Recall**.

Chorus: Wolf, you're getting smarter already.

Wolf: This is fun. What's the next step?

Analysis: That's me. My name is **Analysis**. And if you think **Application** was smart, wait until you try me!

Wolf: What do I do when I analyze?

Chorus: What do you do when you analyze?

Analysis: You take things apart.

Chorus: You take things apart???? But, wolves tear things apart? Shhh!!!!

Wolf: I get it. If I know what all the different parts are, then maybe I'll understand the whole picture better. Am I analyzing if I examine the brick house closely so I could detect all its parts, and, just maybe, I'll find a weak spot? Maybe I'll find an open window or chimney large enough to enter? Or, maybe the bricks in the house are easy to break with a hammer?

Chorus: Ohoooo, a weak spot in the house. Wolf's getting the picture.

Wolf: Okay, I need to go back to the scene. See you later...

Lenny: Just wait one minute....Wolf....you have two more parts to consider. Besides, we need to make sure your brain is really working before you go back to the scene.

Wolf: Well, hurry up....What comes next?

Lenny: The most popular Thinking Skill of all, **Synthesis**.

Wolf: What is **Synthesis**?

Chorus: What is **Synthesis**?

Synthesis: I'm Synthesis, sometimes called **Creativity**. And I'm **much** more complicated than some people think and **much** easier than other people think.

Wolf: **Gee**....You sound more like a riddle.

Chorus: **Synthesis** sounds like a riddle.

Wolf: Gee...that Chorus is giving me a headache. Get rid of them!

(The Chorus members quickly hide their faces.)

Synthesis: **Synthesis** puts things together in a new or different way.

Wolf: You mean I analyze....

Chorus: Take apart.

Wolf: ....so I can synthesize by putting things back together?

Lenny: That's right!

Chorus: What's so hard about synthesizing?

Lenny: Wolf, sometimes it's hard to see a new way of doing things.

Wolf: That doesn't sound very easy to me. But let me try. If I look at all the parts to the whole picture then I can rearrange those parts. Like if I look at the different parts to the third Little Pig's House and then figure out how it's made then maybe I can create **something to use** to help me get into the house or down the chimney?

Chorus: Right....it's figuring out a new way in....to get to the Three Little Pigs...

Lenny: Well, maybe, but then if we synthesize correctly it will lead us to **Evaluation** and maybe to a **different solution** to your problem.

Chorus: A better solution......let's see...

Evaluation: Well, I thought you would never call my name.

Chorus: What will Wolf do when he evaluates?

Handout A3.2,H1, p. 4

Evaluation: He will have to make decisions...sometimes very difficult ones.

Wolf: But, I do that all the time...besides...**hurry up**...I'm getting hungry.

Evaluation: Wolf, you may make decisions all the time...but do you ever stop to consider **WHY** you make those decisions or judgments?

Chorus: Why? Why? Why?

Lenny: And the hardest part is knowing when you need to go back and start over again.

Wolf: You're confusing me. What do you mean?

Chorus: What do you mean?

Lenny: Well, when you are trying to synthesize...

Chorus: Put things together.

Lenny: ....it's not always easy to know if the **new way** is better or worse.

Chorus: But time will tell.

Wolf: Boy, it's sure going to take time for me to learn all this.

Lenny: Wolf, do you think you might need to make a new judgment on **WHY** you want to destroy the Little Pigs?

Chorus: **Destroy the Three Little Pigs?**

Wolf: Well, I don't really know, I have plenty of food to eat. Besides, I really don't like **pig meat!**

Chorus: He doesn't even like **pig meat!**

Lenny: Wolf, maybe you need to exert your energy by trying to help the other Pigs build brick houses instead of trying to figure out how to blow them down. **Just think** what you could accomplish. Maybe the Pigs would even help you build a new brick home for your grandmother; she needs more space.

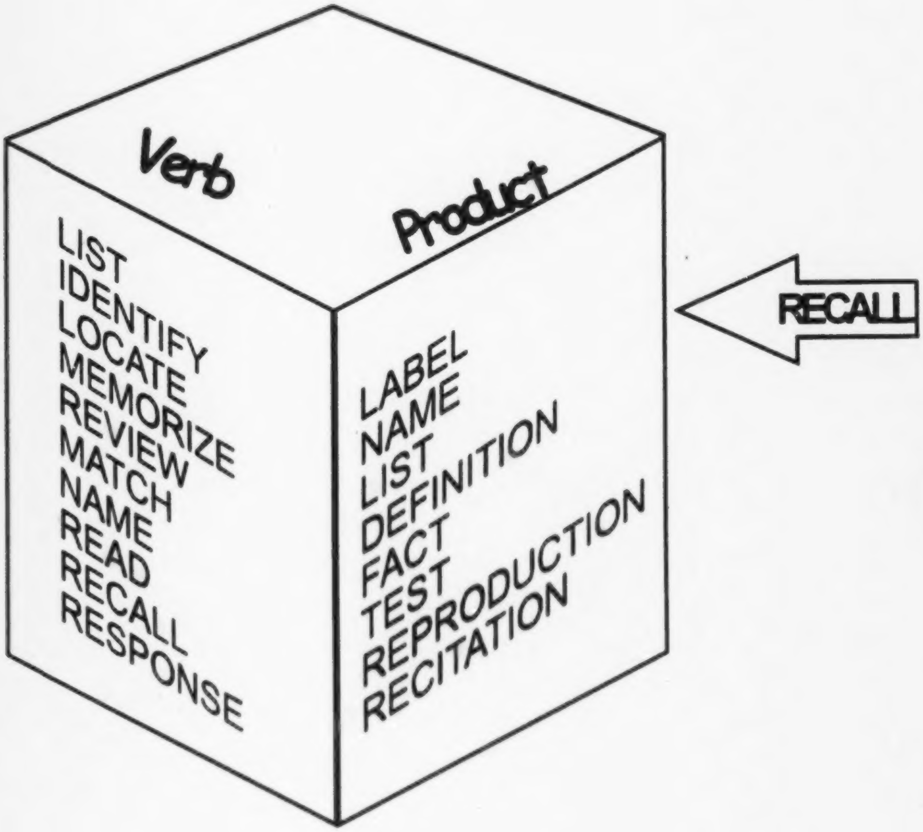
Wolf: Hey, that's really a good idea, I think I like the Blooms! Do you think the Blooms will help work with my cousin? He is hiding in the woods—**waiting for Red Riding Hood**, you know, the cute kid, walking to Grandma's House!

Chorus: Why not? All you need to do is **recall, apply, analyze, synthesize,** and **evaluate**.

Handout A3.2,H1, p. 6

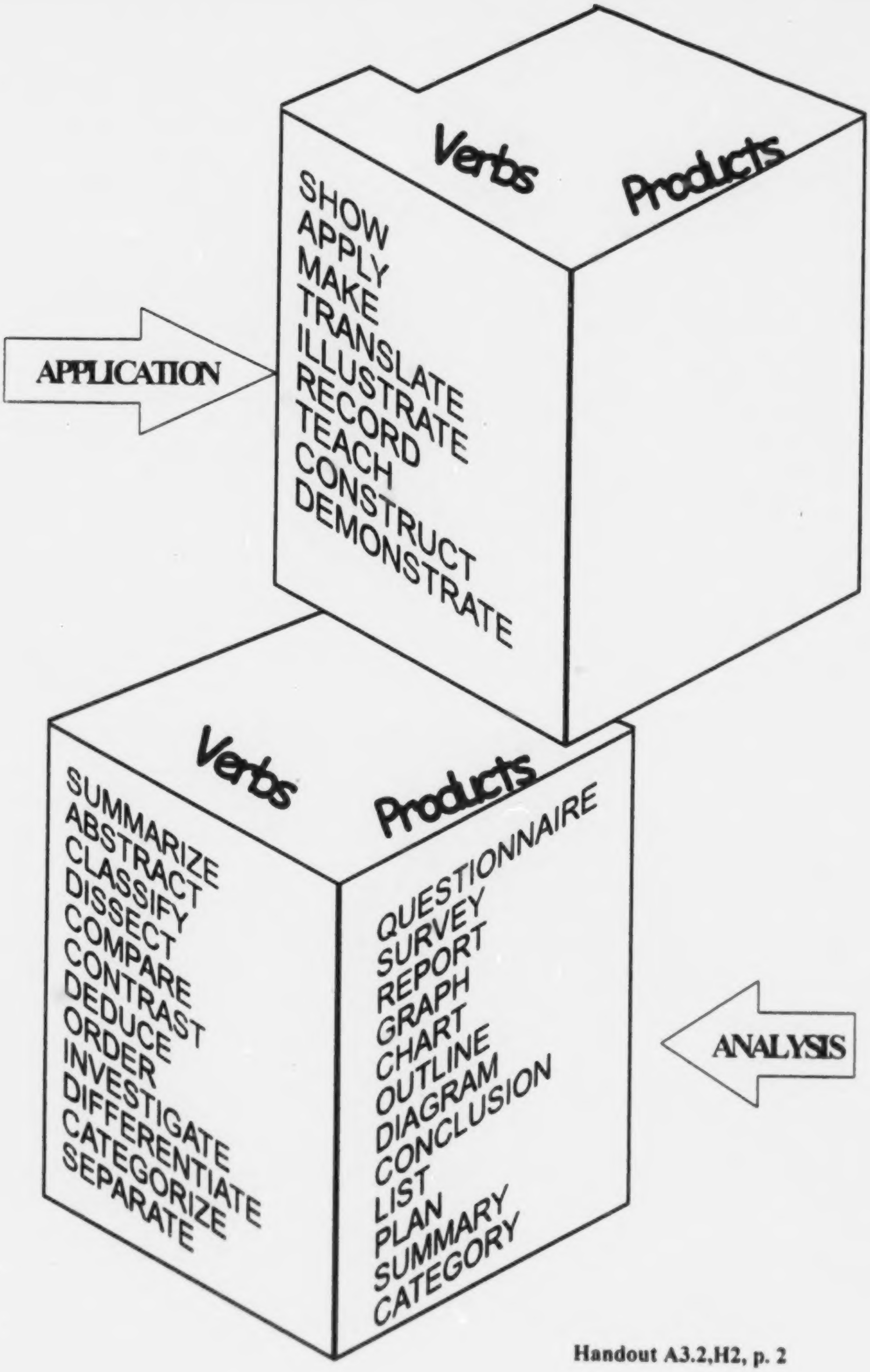
440

Bloom's Taxonomy


Verb
Product
LIST
IDENTIFY
LOCATE
MEMORIZE
REVIEW
MATCH
NAME
READ
RECALL
RESPONSE
LABEL
NAME
LIST
DEFINITION
FACT
TEST
REPRODUCTION
RECITATION
RECALL

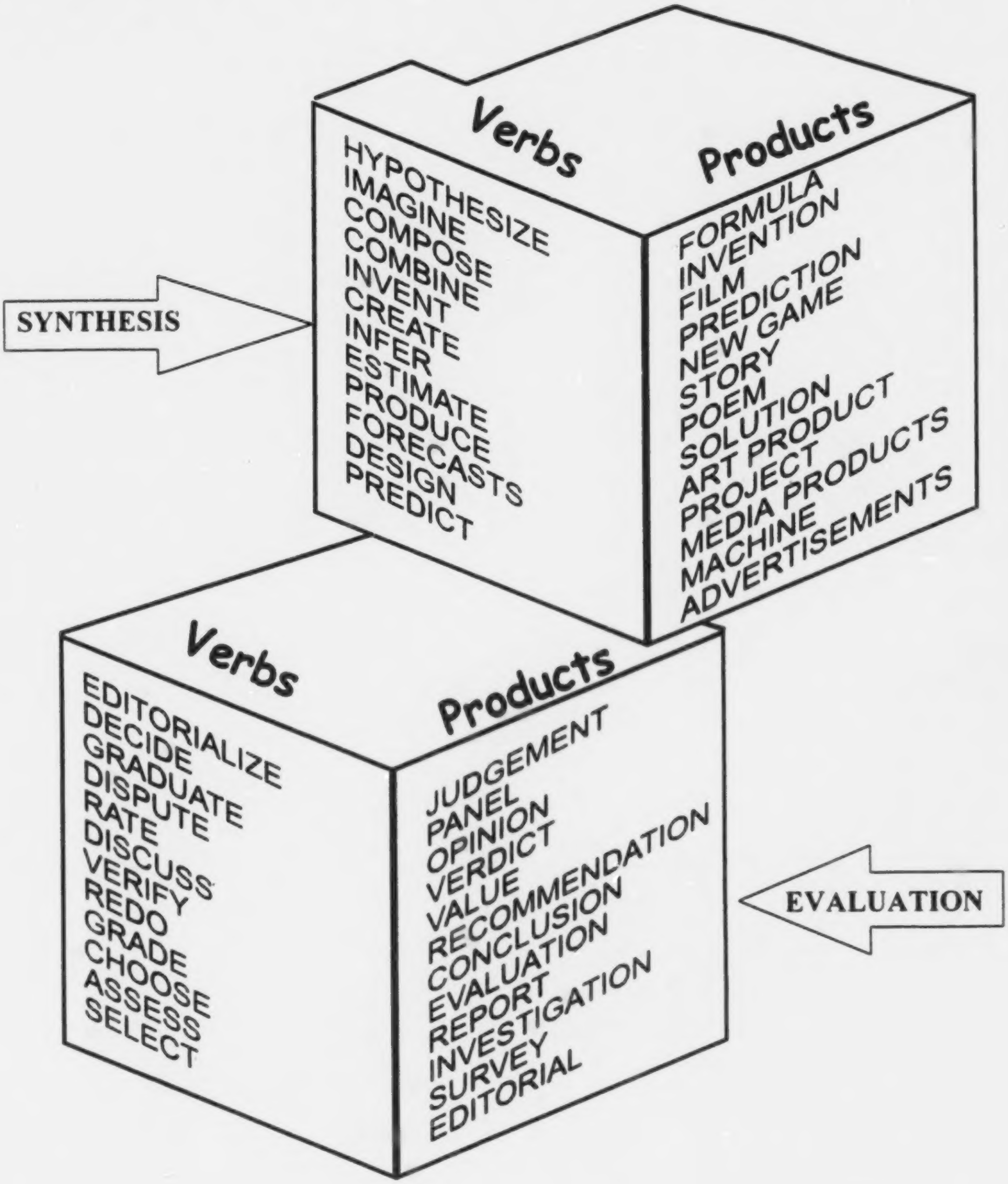
Handout A3.2,H2, p. 1

441


APPLICATION
Verbs
Products
SHOW
APPLY
MAKE
TRANSLATE
ILLUSTRATE
RECORD
TEACH
CONSTRUCT
DEMONSTRATE
Verbs
Products
SUMMARIZE
ABSTRACT
CLASSIFY
DISSECT
COMPARE
CONTRAST
DEDUCE
ORDER
INVESTIGATE
DIFFERENTIATE
CATEGORIZE
SEPARATE
QUESTIONNAIRE
SURVEY
REPORT
GRAPH
CHART
OUTLINE
DIAGRAM
CONCLUSION
LIST
PLAN
SUMMARY
CATEGORY
ANALYSIS

Handout A3.2,H2, p. 2

BEST COPY AVAILABLE


SYNTHESIS
Verbs
HYPOTHESIZE
IMAGINE
COMPOSE
COMBINE
INVENT
CREATE
INFER
ESTIMATE
PRODUCE
FORECASTS
DESIGN
PREDICT
Products
FORMULA
INVENTION
FILM
PREDICTION
NEW GAME
STORY
POEM
SOLUTION
ART PRODUCT
PROJECT
MEDIA PRODUCTS
MACHINE
ADVERTISEMENTS
Verbs
EDITORIALIZE
DECIDE
GRADUATE
DISPUTE
RATE
DISCUSS
VERIFY
REDO
GRADE
CHOOSE
ASSESS
SELECT
Products
JUDGEMENT
PANEL
OPINION
VERDICT
VALUE
RECOMMENDATION
CONCLUSION
EVALUATION
REPORT
INVESTIGATION
INVEY
SURVEY
EDITORIAL
EVALUATION

Handout A3.2,H2, p. 3

Using the Question (Assessment Task) Quilt with Students

1. Have students review Bloom's Taxonomy using the Choral reading in class.
2. After a lesson or video or at the end of a unit of study, allow students to create their own questions or other assessment tasks to place on a Question (Assessment Task) Quilt. Different groups could be given different levels of Bloom's Taxonomy on which to focus their questions or tasks.
3. The teacher can pull questions from the quilt for tests, quizzes, etc. This approach should encourage the students to take more responsibility for their own learning. It allows a teacher to integrate assessment and instruction by using the assessment to teach the very skills they also want to assess.

Example: Questions and tasks created and posted by students for a particular unit of study

Bloom's Level	Questions Week 1	Questions Week 2	Questions Week 3	Questions Week 4	Questions Week 5
Recall	Label the parts of the human cell.	Define in writing the word mitochondria.	Name three functions of the nucleus of an animal cell.	Who coined the term cell?	Match the structure with the function of the listed parts.
Application	Categorize the cell structures as plant, animal or both.	?	?	?	?
Analysis	Compare and contrast plant and animal cells.	?	?	?	?
Synthesis	Using the graph, predict the influence of salt water on the celery (plant) cells.	?	?	?	?
Evaluation	Argue that cell size is limited and that number of cells determine size of animals.	?	?	?	?

Handout A3.2,H3

444

ENGLISH QUESTIONS//HIGH SCHOOL (Cut out sections and place in envelopes)

A	B	C
Choose the correct punctuation for this sentence: Does John play basketball __ a) . b) ! c) ? d) ,	Copy the sentences in exercise A from your grammar workbook onto a sheet of notebook paper. Punctuate them correctly.	You have been asked to proofread your older brother's book report. He wants you to check the punctuation and will pay you $10.00 for your services. Using a highlighter, mark the errors you find and correct them.
D	**E**	**F**
Write an entry in your journal using *no* ending punctuation. After you have finished writing, get into a group with two other people. Exchange journals and try to read each other's papers. After you have read aloud, write a new journal entry about what you learned about the purpose/need for ending punctuation.	You are entering a short story contest sponsored by your school. After you have written a rough draft, proofread it for good punctuation skills. After carefully checking for errors, re-write your story making all corrections. Following contest guidelines, prepare your corrected copy for submission to the contest.	Make a list of as many jobs as you can think of where good punctuation skills are necessary to be successful. Ex: newspaper reporter. After making your list, pick four jobs from the list and write a paragraph on each, explaining *why* these jobs required good punctuation skills. (Make sure you use good punctuation skills when writing your report.)
G	**H**	**I**
Create 10 sentences of your own representing the four types of basic sentences: declarative, interrogative, exclamatory, and imperative. Make sure you punctuate these sentences correctly.	Pick out any three of your journal entries over the past few weeks. Analyze/count the types of sentences you used most often in your writing. Make a punctuation graph to show your findings. Now write a new journal entry about what you learned from creating this graph. (Ex: I need to use more sentence variety in my writing.) Incorporate what you learned into your newest journal entry.	Explain how punctuation marks in writing are like road signs on the highway.

Handout A3.2,H4

SCIENCE QUESTIONS//ELEMENTARY/MIDDLE SCHOOL (Cut out sections and place in envelopes)

<table>
<tr><th>A</th><th>B</th><th>C</th></tr>
<tr><td>If you could eliminate any organ in the body, which one would it be and why? What complications might arise without this organ?</td><td>Make a diagram and tell the function of each organ of any of the following systems: respiratory, digestive, circulatory, or nervous.</td><td>Make a diorama of an extremely unhealthy environment that the respiratory system should NOT be exposed to.</td></tr>
<tr><th>D</th><th>E</th><th>F</th></tr>
<tr><td>Using the attached drawing, identify the bones of a skeleton and label them.</td><td>Imagine that you are a cookie about to be eaten. Write a travel diary as you travel through the digestive system.</td><td>The heart is an organ of which of the following systems?
a) nervous
b) respiratory
c) circulatory
d) digestive</td></tr>
<tr><th>G</th><th colspan="2">H</th></tr>
<tr><td>Using a diagram to show the body location, write in complete sentences an explanation which will compare and contrast the appendix to the gall bladder.</td><td colspan="2">Sally, one your classmates, has diabetes. You know she has a difficult time regulating her diet at school. Discuss with your peers some of the problems Sally has related to her diabetes. Make a recommendation for a daily meal and snack plan for Sally to follow while at school and offer her a way to help monitor her progress.</td></tr>
<tr><th colspan="3">I</th></tr>
<tr><td colspan="3">Show where your "funny bone" is located and report why it feels so strange when you hit it.</td></tr>
</table>

Handout A3.2,H5

SOCIAL STUDIES QUESTIONS//HIGH SCHOOL (Cut out sections and place in envelopes)

A	B	C
How many main branches are there in the U.S. federal government? a) 1 b) 2 c) 3 d) 4	The U. S. government has three major branches: executive, legislative, judicial. Write a definition for each of these. You may use your textbook.	Pretend that you are a member of the U. S. House of Representatives. You want to introduce a bill that would raise the minimum wage. Design a flow chart that shows the steps you must go through to get the bill introduced.
D	**E**	**F**
Not all governments are structured like the U. S. federal government. Choose a foreign country and compare and contrast its national government structure with the three branch U. S. system.	The following activities are governed by local/state/or federal regulations. How old must you be to legally: Vote? Drive a car? Run for governor? Become a police officer? Get arrested? Buy cigarettes? Work a 40-hour week?	You have been falsely accused and convicted of a felony. You and your lawyer have vowed to fight this miscarriage of justice to the highest court in the land—the Supreme Court. You must understand the justice system to make it work for you. Outline the process you will have to follow to win your freedom. What's the next step you must take? Then the next? Follow the process all the way to the Supreme Court. (Good luck!)
F	**G**	**H**
The three branches of government provide checks and balances for each other. List three of these checks and balances and explain how they work	List the basic qualifications a person must have to be able to run for President of the U. S. Now make a second list of qualifications you consider necessary to become Chief Officer of the Executive branch of our government. Be prepared to explain why.	If you had to eliminate any one of the three branches of government, which one would it be? What are some of the problems that might arise if this branch were eliminated?

Handout A3.2,H6

SOCIAL STUDIES QUESTIONS//ELEMENTARY (Cut out sections and place in envelopes)

A	B	C
Make a model of a community. Make sure to include important services, resources, and buildings found in a community.	Imagine that you could design your own community. What would it look like? Describe the services and resources that might be found in your community.	The leader of a community is called the: a) president b) governor c) major d) vice president
D	**E**	**F**
A classmate has just moved into your community. List important places that she should visit and important services that she should know about.	Design a travel brochure that describes your community.	Using the attached map, label the following parts of a community: hospital, school, police station, fire station, park.
G	**H**	**I**
Make a cartoon strip that tells several facts about your community.	Tell why it is important to have community services such as policemen and firemen.	Why is it important for us to protect our resources within our community?

Handout A3.2,H7

SOCIAL STUDIES//ELEMENTARY—BLANK (Cut out sections and place in envelopes)

<table>
<tr><td>A</td><td>B</td><td>C</td></tr>
<tr><td>D</td><td>E</td><td>F</td></tr>
<tr><td>G</td><td colspan="2">H</td></tr>
<tr><td colspan="3">I</td></tr>
</table>

Handout A3.2,H8

Activity 3.3

Performance Criteria—Keys to Success

Purposes:

1. To learn the characteristics of sound performance criteria—what do good performance criteria look like?
2. To learn the additional features of performance criteria that make them most useful as instructional tools in the classroom
3. To summarize previous points about the relationship of assessment purpose to design of performance criteria
4. To build skill in reviewing and critiquing performance criteria
5. To practice one method of teaching criteria to students

Uses:

This is an advanced level activity that can be used in **Chapter 3** to explore the characteristics of quality performance criteria (rubrics) and discuss the advantages and disadvantages of different kinds of performance criteria. It can also be used to follow-up **Activity 1.9**—*Going to School* and **Activity 2.1**—*Sorting Student Work.*

Prerequisites might include (a) an activity on the rationale for alternative assessment (e.g., **Activities 1.1**—*Changing Assessment Practices...*, **1.6**—*Comparing Multiple-Choice and Alternative Assessment*, or **1.12**—*Assessment Principles*); (b) activities that build knowledge of the role of performance criteria in a performance assessment, how to use criteria to score student work, and how to use criteria as instructional tools in the classroom such as **Chapter 2** text and **Activities 2.1**—*Sorting Student Work*, and **2.5**—*How Knowledge of Performance*

Criteria Affect Performance; and (c) one of the activities providing a gentle introduction to quality issues, such as **1.5**—*Clapping Hands* or **1.11**—*Assessment Principles*.

Rationale:

A variety of performance criteria types are being developed. The type to use depends on: (1) the type of student outcome being assessed; (2) whether the assessment is large-scale or classroom; and (3) the vision one has of the role of assessment in the educational process. These issues are discussed in the text in **Chapter 3**. The purpose of this exercise is to demonstrate, with specific examples, the characteristics of high quality performance criteria *if the purpose is to maximize classroom instructional usefulness*. At the end of the activity we want participants to feel confident that they would know how to reply to a colleague who comes up to them and asks, "I'm thinking about using this set of performance criteria in my classroom. What do you think about them?"

The activity also models the process for teaching criteria to students. When teachers teach criteria they need to define each trait, give some examples, have students practice identifying strong and weak examples of products/performances using the criteria, and have students practice making weak products/performances stronger. While students might be working on, say, criteria for quality oral presentations, participants in this professional development activity are working on *criteria for criteria*. Thus, in this activity, we define criteria for criteria and identify strong and weak examples of criteria. You can even make this into a little assessment joke because the criteria for criteria have to adhere to their own standards—content/coverage, clarity, and generality.

Materials:

For both options:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads: A3.3,O1**—*Performance Criteria Keys Purposes*; **A3.3,O2**—*Performance Criteria—Keys to Success*; **A3.3,O13**—*Speaking Criteria to Critique*; **A3.3,O14**—*Advantages and Disadvantages of Task Specific and General Criteria*; **A3.3,O15**—*Advantages and Disadvantages of Holistic and Analytical Trait Criteria*; **A3.3,O16**—*Informal Writing Inventory*.

In addition, choose *one* of the following:

Option 1: Mathematics

- **Overheads:** Two "name the graph" student solutions, **A3.3,O3**—*Number of Teeth;* **A3.3,O4**—*Fruits;* **A3.3,O5**—*Skimpy Criteria*; **A3.3,O6**—*Task Specific Criteria, Version*

1: Name the Graph and **A3.3,O7**—*Task Specific Criteria, Version 2: Name the Graph* (these can also be used as handouts)

- **Handouts:** Several rubrics from **Appendix A**—*Sampler* are required at various steps in this activity. There is a special part of the **Appendix A**—*Sampler Index* devoted to **Activity 3.3** to help users locate appropriate mathematics rubrics.

Option 2: Writing

- **Overheads:** Two writing samples, **A3.3,O8**—*Redwoods* and **A3.3,O9**—*Fox;* **A3.3,O10**—*Skimpy Writing Assessment Criteria*; **A3.3,O11**—*Sample Holistic Writing Criteria*; and **A3.3,O12**—*Sample Task specific Writing Criteria* (these can also be used as handouts)
- **Handouts:** Several rubrics from **Appendix A**—*Sampler* are required at various steps in this activity. There is a special part of the **Appendix A**—*Sampler Index* devoted to **Activity 3.3** to help users locate appropriate writing rubrics.

Time Required:

60-90 minutes

Facilitator's Notes:

A. (2 minutes) **Introduction**. Use **Overhead A3.3,O1**—*Performance Criteria Keys Purposes* to introduce the activity. The facilitator's notes include two running examples—math and writing. You will likely only be using one of the two in any single setting. Directions that relate to both activities are not boxed. Boxed instructions mean that you need to choose either the math or writing version.

B. (20 minutes) **Key 1: Coverage**

1. The first "key to success" (see **Overhead A3.3,O2**—*Performance Criteria, Keys to Success*) for performance criteria is: Don't leave out anything of importance. Describe why it is essential to cover everything of importance and give examples. You might say something like: *Well-defined performance criteria are a statement of what is valued in student work. They communicate to us, as teachers, and students what to do and what is important to attend to. If important dimensions of performance are left out, it is likely that they won't be emphasized in instruction or student work because they won't be assessed. An example in writing is a state that left the trait of "voice" out of their rubric because they felt it would be too difficult to get good rater agreement. The result—no systematic attention to voice in writing. Another example is in music. What if you were a fiddle player and bowing wasn't rated as part of performance? What message does that communicate? Or, if teachers were held*

accountable for student performance, what effect might leaving bowing out have on instruction? A third example might be the Olympics. What if entry into the water was not taken into account in diving? Instead, type of bathing suit is scored. What message does this send about what is important? How would it affect coaching?

Ask participants to think of examples from their own experience about how criteria communicate what counts. Think: driver's test, Olympics, what usually gets rewarded in student writing, etc.

2. Choose either math or writing.

Math: Show the strong student performance (**A3.3,O3**—*Number of Teeth*). Ask, *How well did this student communicate what he or she did?* Most teachers agree that the student did a good job on this trait. Now put up the *Skimpy Criteria* (**A3.3,O5**) and point out the section on communication. Then cover up the trait of communication and ask, *What message would be sent if communication was not scored? How might instruction be affected? How might student work be affected?*	**Writing:** Show the strong student performance (**A3.3,O9**—*Fox)* and give participants the *Skimpy Writing Criteria* (**A3.3,10**). Ask, *How well did this student use detail to add clarity?* Most teachers agree that the student did a good job on this trait. Then cover up the trait of detail and ask, *What message would be sent if detail was not scored? How might instruction be affected? How might student work be affected?*

3. Besides covering *all* those aspects of a performance that are truly aspects of a strong performance, it is equally important to *leave out* those aspects of performance that do not truly distinguish quality. A great example is the "two minute speech." *The son of a friend had to make an oral presentation as part of a seventh grade culminating project for the semester. The speech had to be between two and three minutes long. He practiced at home and was timed at two minutes 15 seconds. When he gave his oral presentation at school he was a little nervous and was two seconds under two minutes. His grade went from an 'A' to a 'C.'* What message does this send to students about what is important?

4. In order for participants to really understand the importance of good content coverage, it is useful to practice finding and correcting the "holes" in criteria/rubrics. Overhead **A3.3,O13**—*Speaking Criteria to Critique* can be used for this. Ask participants to work in groups to identify important aspects of performance that have been left out or not covered completely.

   (What other groups have said: There should be a trait for "content." Under delivery one might add hand gestures and posture. Students also need to understand that oral presentation is organized and presented different than written presentation—sentences

453

are shorter, there is more repetition, and fewer points can be made in the same amount of time. Where should multi-media and visuals go? Where should personal style go?)

Now rate **A3.3,O13**—*Speaking Criteria to Critique* on a scale of 1-5 where 1 denotes "lots of holes; coverage is weak" and 5 denotes "not many holes; coverage is good. (These speaking criteria usually get a 3-4 rating on the trait of coverage.)

5. **A3.3,O16**—*Informal Writing Inventory* is usually rated very weak on the trait of coverage, so it provides a good example to anchor the low end of the scale. The Informal Writing Inventory (IWI) was developed to screen students into special education classes. The purpose is to identify whether there is a writing disability. Students are shown a picture and write about what is happening in the picture. This is a nice writing task; pictures are frequently used to prompt primary writing. The criteria, however, are these: First one counts up the number of grammatical errors. Then one counts up the number of fatal grammatical errors—those that block one's ability to understand what the writer means. Then one forms two ratios—the error index which is the ratio of erroneous writing to well-formed writing; and the communication index, which is the percentage of total errors that disrupt communication.

   Participants usually feel that counting grammatical errors is not a very good criterion for good writing. It leaves out a lot that is important. Therefore, the *coverage* is rated very low. (It anchors the "1" end of the scale.)

6. Be sure to emphasize the main point here—teachers teach to the criteria and students learn to the criteria. Therefore, criteria/rubrics must cover that which most distinguishes quality, and they must leave out those things which are not really indicators of quality. (If things like following directions, having a date on the paper, etc. are important outcomes, score them separately. Don't mix "following directions" with the quality of the student response.)

7. If you would like to give participants more practice on critiquing performance criteria for "contact/coverage," see the **Activity 3.3 Index** part of **Appendix A**—*Sampler*.

C. (20 minutes) **Key 2: Detail**

1. The second key to success (**A3.3,O2**) is having criteria that are clear, detailed, and illustrated with samples of student work.

Math: Refer to the *Skimpy Criteria* (**A3.3,O5**). Have participants score the strong and/or weak student performance (**A3.3,O3** or **A3.3,O4**) on the trait of *communication*. Chances are that rater agreement will be pretty good. Now, pose the dilemma: *What if you give feedback to a student using this scale and the student says, "I know I communicated well because my score is high, but I don't know why. Why did I get a good score on communication?"*	**Writing:** Refer to the *Skimpy Criteria* (**A3.3,O10**). Ask participants to score the strong and/or weak student performances (**A3.3,O8** or**A3.3,O9**) on the trait of "personal expression." Chances are that rater agreement will be pretty good. (*Fox* is very high and *Redwoods* is very low.) Now pose the dilemma: *What if you give feedback to a student using this scale and the student says, "I know I exhibited a sense of personal expression because my score is high, but I don't know why. Why did I get a good score on personal expression?*

The point to make is that students have a harder time generating another strong response if they don't know what made the previous response strong. Or if the previous response is weak, they have a harder time making the next one better if they don't know what made the current one weak. *The point of having clarity and detail in criteria is not necessarily to ensure that two different raters would give the student work the same score (rater agreement) although detail will help here too; rather, detail helps to clearly communicate to students what they need to do to produce quality work.*

Also note that criteria can have good content/coverage and still be weak in clarity. For example, either of the *Skimpy Criteria* covers the important dimensions of performance; it is just not clear what many of the terms mean, or what the difference is between, for example, a '1' and a '2.'

2. Ask participants to rate *Skimpy Criteria* (**A3.3,O10** or **A3.3O5)** for clarity on a scale of one to five where '1' is low and '5' is high. Remind them to base their rating on two questions: (a) are these clear enough that two different raters are likely to give the same score to the same piece of student work; and (b) would these criteria communicate to a student who is not doing well what to do differently next time? (Both *Skimpy Criteria* examples usually receive a low score for clarity.) Ask participants to outline what they would do to make the criteria clearer. (Responses

generally include: add definitions, describe indicators of 1-5 performance; find work samples that illustrate what you mean.)

Now ask participants to rate the detailed analytical trait criteria selected from the **Appendix A**—*Sampler* on the trait of 'clarity' where '1' is low and '5' is high. (These rubrics get a higher score, usually around '4' for math and '5' for writing) Ask participants to outline what they would do to make the criteria even clearer.

3. Make sure that participants understand the relationship of purpose to design. If the purpose of the criteria is use as an instructional tool in the classroom, then adequate detail is of paramount importance. If the purpose is merely to put numbers on student work for large-scale assessment, perhaps not as much detail is important.

4. Ask participants to find another example of skimpy criteria in the *Sampler* (**Appendix A**) and an example of good detail and moderate detail. (See the **Appendix A**—*Sampler* **Activity 3.3 Index** for suggestions.)

D. (20 minutes) **Key 3: General**

1. The third key to success (**A3.3,O2**) is using generalized performance criteria. Define "generalized" and "task specific" and discuss examples. You could say: *"Generalized" means that the criteria can be used to rate a whole class or category of performances and not just the work from a specific task. Task-specific criteria can only be used to rate performance on a single task. For example, in the Olympics the same criteria are used to rate all dives—difficulty, form, entry into the water, etc.; there are not separate criteria for each type of dive. In writing assessment, criteria are often very general—some places use the same criteria to rate all kinds of writing; others have separate rubrics for persuasive, narrative, and expository writing. But, very few are task-specific—what sense would it make to have a separate rubric for each prompt?*

Math: Illustrate task-specific criteria with **A3.3,O6**—*Task Specific Criteria, Version 1: Name the Graph*, and **A3.3,O7**—*Task Specific Criteria, Version 2: Name the Graph*. Ask participants whether they could use these criteria to judge student performance on another task (take any other math task from the *Sampler*. (The answer is "no." This is the essence of task-specific scoring; it applies only to a single task.) Then ask whether the general criteria (selected from the **Appendix A**—*Sampler* for this purpose) could be used to judge student performance on *Name the Graph*. (The answer is "yes." This is the essence of generalized criteria; they apply across tasks.)	**Writing:** Illustrate task specific criteria with **A3.3,O12**—*Task Specific Writing Criteria*. Ask participants whether they could use these criteria to judge student performance on another task (say, "write a persuasive essay on school uniforms"). (The answer is "no." This is the essence of task-specific scoring; it applies only to a single task.) Then ask whether the general criteria (selected from the **Appendix A**—*Sampler* for this purpose) could be used to judge student performance on *Describe a Favorite Place*. (The answer is "yes." This is the essence of generalized criteria; they apply across tasks.)

2. To make sure participants know the difference between task specific and general criteria, ask them to find examples in the **Appendix A**—*Sampler*. (The *General Index* lists which samples illustrate the two kinds of criteria.)

3. Now discuss the relative advantages and disadvantages of these types of criteria. (**Activity 1.9**—*Going to School* also includes such a discussion.) As a guide to or summary of the discussion use **A3.3,O14**—*Advantages and Disadvantages of Task Specific and General Criteria.*

Advantages of task-specific criteria:

a. Quicker to train raters.

b. High rater agreement right from the start.

c. Therefore, often used in large-scale assessment.

Disadvantages of task-specific criteria:

a. Have to develop new criteria for each task.

b. Makes no sense to show them to students ahead of time because they "give away" the answer.

c. Can't use them for judging the work in portfolios because the content of each student's portfolio can be very different. (This is especially true at the large-scale level.)

d. What happens if a student comes up with a perfectly reasonable strategy or solution, but it isn't one included in the task-specific criteria? This happens frequently in large-scale assessment when raters are going fast and not really thinking about what they are doing. In fact a personal communication from a major test publisher indicated that this was a major reason they were moving away from task-specific scoring.

e. Related to (d) above—task-specific criteria *do not make the rater think*, the thinking has already been done for the rater. The developer of the rubric already thought through, for example, how good problem solving would look on *this* problem. Therefore, *the rater* doesn't have to think. Likewise such criteria don't make students think. If a major reason for developing criteria is as a tool for learning in the classroom (e.g., students learn standards of quality for work), then generalized criteria do a better job because we want students to think—to be able to generalize what they learned on one task to make performance on the next task better.

(A sample script for the last point might be: *The value of generalized criteria is that they help us to define what "good" looks like so that we can begin to generalize across performances and bring information from past experiences to bear on new* *experiences. This is especially important for the "hard to define" or "disagreement on what this means" goals such as critical thinking, problem solving, collaborative working, communication skills, etc. If people already agree on the definitions, then you can cut corners. But, people often* *don't* *agree on what some skills mean and what it looks like when students do "it" well. We can't hold students accountable for different visions of the same target. It is our moral obligation to define precisely what we mean.)*

Advantages of generalized criteria:

a. They help students generalize what quality looks like from one task to the next.

b. It is very beneficial to ask students to help develop them.

Disadvantages of generalized criteria:

a. They take longer to learn (because the rater has to learn how to apply them to a variety of tasks).

b. Rater agreement, at least at first, is lower.

4. We don't necessarily mean to imply that task-specific criteria should never be used. Remember, rubric design relates to purpose—what they will be used for. It might be, for example, that task-specific criteria are better used when one wants to determine the student's *conceptual understanding* of specific aspects of a task or if you want the student to use specific equations or come up with specific substeps or final answers. Generalized criteria might be best for "big" outcomes such as problem solving, communication in math, group collaboration, and critical thinking—process skills that *should be* generalized across tasks. They are also, obviously, necessary when you will be looking at work generated by students on different tasks as, for example, in portfolios.

   A compromise might be to "mix and match" generalized criteria and task-specific criteria. For example, one might, on a task, overlay generalized criteria for group collaboration and critical thinking, and also have task-specific criteria for specific substeps. Another compromise might be to have students *consciously* develop task-specific criteria from a generalized rubric. That way they practice the ability to generalize.

E. (20 minutes) **Key 4: Analytical Trait Criteria**

1. We hesitated before putting this on this list of quality criteria because choice of analytical trait or holistic really depends on purpose. For some uses, such as large-scale summative assessment, where speed of scoring is an issue, holistic scoring may be a fine choice. However, for classroom instructional uses, analytical trait criteria are more useful. Be sure participants are clear about this. We don't want to create

analytical trait zealots that write nasty letters to their state departments of education for using holistic rubrics. So...

If the purpose for criteria is using them with students to promote learning, the fourth key to success (**A3.3,O2**) is analytical trait rather than holistic criteria. Hopefully you have already done activities that illustrate the difference and why it is important (such as **Activity 2.1**—*Sorting Student Work).* In a pinch, you might use the following script. (But, telling is never as effective as showing.) You might say, *Analytical trait means that more than one dimension of performance is assessed. Holistic means that you make one overall judgment of the quality of the response.*

2. Have participants score the student performances with the analytical trait and holistic rubrics selected from the **Appendix A**—*Sampler* for this activity. (Suggestions are provided in the **Activity 3.3** part of the *Sampler Index*. A sample holistic writing rubric is attached as **Handout A3.3,O11**—*Sample Holistic Scoring Criteria.*) Discuss advantages and disadvantages of each method (see **A3.3,O15**—*Advantages and Disadvantages of Holistic and Analytical Trait Criteria).*

   The bottom line is that analytical trait systems are not worth the effort in the classroom if all they are to be used for is putting grades on student papers. If, however, they are used as an instructional methodology—to focus instruction, communicate with students, allow for student self-evaluation, and direct instruction on traits—they are very powerful. In short, purpose dictates design.

3. To consolidate the notion of holistic and analytical trait criteria, ask participants to browse through the **Appendix A**—*Sampler* and find examples of each type. (The general index to **Appendix A** will help you locate relevant samples.)

F. (10 minutes) **Wrap-Up**. Make a game out of repeating the mantra for quality criteria—coverage, clarity, generality (and maybe analytical trait if the purpose is instructional). Ask participants to repeat this mantra to each other.

Ask participants to explain to each other why these features of criteria are important for the purpose of classroom use in instruction. (And, if the purpose were different, say, large-scale screening of students, what rubric features would be important.)

Then ask participants to write in their own words what it means to have quality criteria. If a colleague were to come up to them and ask, "I'm thinking about using these criteria in my classroom, what do you think?" what would they look for?

Performance Criteria—Keys to Success

Purposes:

1. To learn characteristics of high quality performance criteria
2. To learn features of performance criteria that make them useful as instructional tools in the classroom
3. To build skill in reviewing and critiquing performance criteria

Overhead A3.3,O1

Performance Criteria—Keys to Success

1. Content/Coverage:

Cover everything of importance—don't leave things out. What they see is what you'll get. Could you explain to someone else why each thing assessed is essential? What you left out and why? **Only include the things of most importance.** It might be desirable to have a speech that is only two minutes long, but is this really what is most important in an oral presentation?

Overhead A3.3,O2,p. 1

461

Performance Criteria—Keys to Success

2. Clarity:

The more detail the better. Write definitions; use adjectives and descriptive phrases. Could you explain to someone else what each item in the criteria means? **Have samples of student work.** Illustrate what you mean. Could you show someone else sample student work or performances that specifically illustrate each score point on each dimension?

3. Generality:

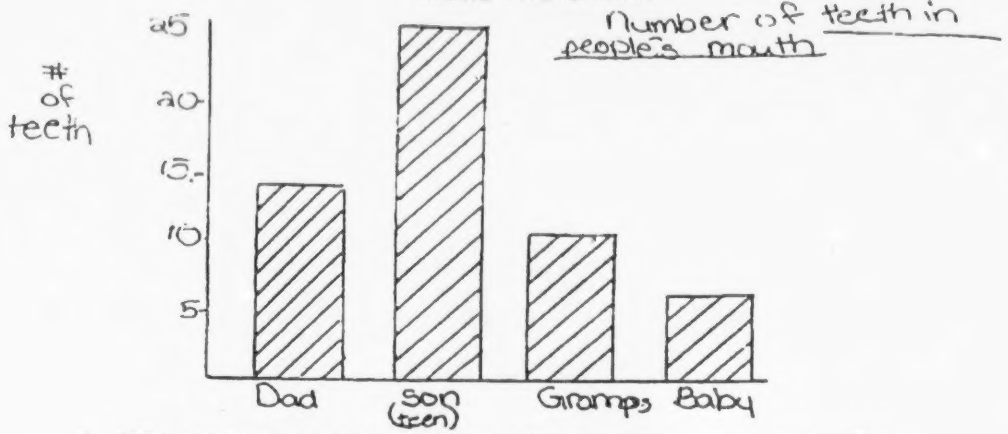
Aim for generality. Develop criteria that can be used across tasks. Could you show someone else samples of strong and weak performance from many different assignments and tasks? And explain how the criteria relate to these different situations?

4. Consider analytical-trait (several ratings) rather than holistic (one rating) for classroom instructional uses

Overhead A3.3,O2, p.2

Graphing Problem - Grade 5

NAME THE GRAPH



1. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.

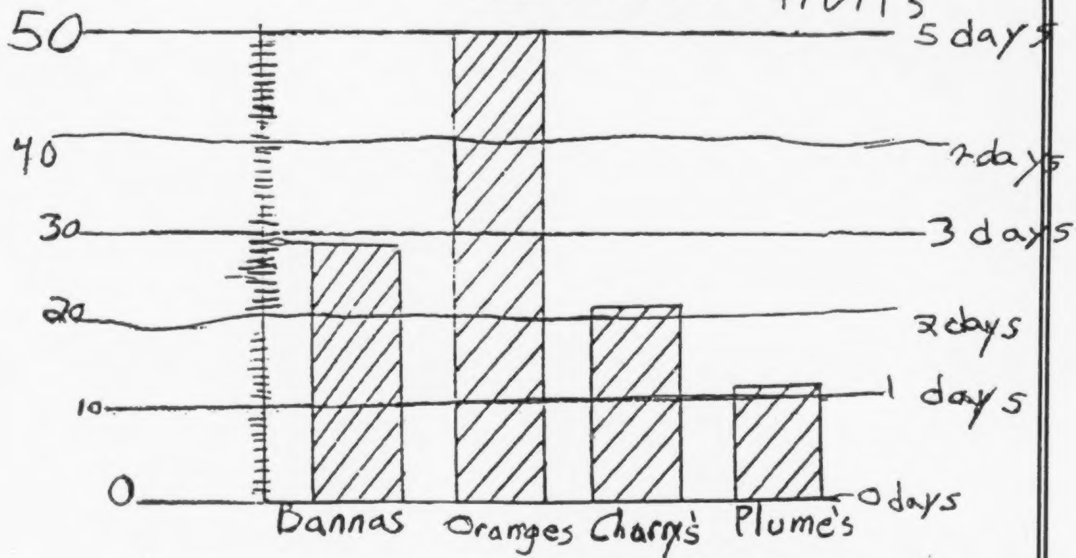
2. Write down everything you know from your graph.

1. Son has more teeth
2. Baby has the least
3. Gramps has more teeth than Baby.
4. Dad has more teeth than Gramps & Baby.
5. Son has about 25 teeth.
6. Dad has about 15 teeth.
7. Gramps has about 10 teeth.
8. Baby has about 6 teeth.
9. The name of the graph is # of teeth in people's mouths
10. The graph is about Baby, Gramps, Dad, and Son.
11. On this graph the most teeth you can have is 25.
12. Son has more teeth - than Baby.

Overhead A3.3,O3

BEST COPY AVAILABLE

NAME THE GRAPH fruits



1. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.

2. Write down everything you know from your graph.

One day Mike ate 12 plumes. The next day Mike ate 21 charrys. The next day Mike ate 29 Banna's. The next day Mike ate 5 oranges. If you add them up he ate 112 In all.

Overhead A3.3,O4

BEST COPY AVAILABLE

Skimpy Criteria

	Frequently	Occasionally	Sometimes	Never
Understanding the problem	4	3	2	1
Appropriate use of math	4	3	2	1
Clearly focused, good reasoning	4	3	2	1
Effective communication	4	3	2	1

Overhead A3.3,O5

Task-Specific Criteria, Version 1: Name the Graph

4: Has a title under each bar, a label for each axes, and appropriate values listed on the y-axis. Has a title for the whole graph. The values along the y-axis are spaced appropriately and relate to the relative heights of the bars. The labels under each bar make sense. There are at least five things listed for the second question.

3: The student graph is essentially correct, but has at least one major thing wrong. For the most part, conclusions in part 2 are reasonable—at least three of the five statements are correct.

2: The student shows some understanding of the sorts of things that might go into a graph, such as the need for labels and numbers, but the response has serious errors, such as: labels that don't make sense; and numbers on the y-axis that don't bear a relationship to the height of the graphs. Conclusions are sketchy.

1: Most parts of the graph display lack of understanding, are missing, or are just plain wrong. Conclusions do not logically follow from the graph.

Overhead A3.3,O6

Task Specific Criteria, Version 2: Name the Graph

Number of Points	Feature
1	Graph titled
1	x-axis labeled
1	y-axis labeled
1	Bars labeled
2	Values on y-axis spaced appropriately and related correctly to bars
3	1 point for up to two correct conclusions 2 points for three to four correct conclusions 3 points for five or more correct conclusions

Overhead A3.3,O7

The Redwoods

Last year, we went on a vacation and we had a wonderful time. The weather was sunny and warm and there was lots to do, so we were never bored.

My parents visited friends and took pictures for their friends back home. My brother and I swam and also hiked in the woods. When we got tired of that, we just ate and had a wonderful time.

It was exciting and fun to be together as a family and to do things together. I love my family and this is a time that I will remember for a long time. I hope we will go back again next year for more fun and an even better time than we had this year.

Overhead A3.3,O8

FOX

I don't get along with people to good, and sometimes I am alone for a long time. When I am alone, I like to walk to forests and places where only me and the animals are. My best friend is God, but when I don't believe he's around sometime's, my dog stands in. We do every thing together. Hunt, fish, walk, eat and sleep together. My dog's name is Fox, 'cause he looks like an Arctic Fox. Fox and I used to live in this house with a pond behind. That pond was our property. The only thing allowed on it (that we allowed) was ducks & fish. If another person or dog would even look like going near that place, Fox and I would run them off in a frenzy. There was a lot of rocks around, so I would build forts and traps for any body even daring to come near. The pond had a bridge that was shaded by willows, so on a hot day me and Fox would sit on that bridge & soak our feet, well, I would soak my feet, Fox just kinda jumped in.

Overhead A3.3,O9, p. 1

At night, the pond was alive with frogs, so I would invite this kid over, (he was a guy like me) and catch frogs. After we had a couple each, we would pick the best looking one out of our group and race them. The winner gets the other guys frog.

In the winter, the pond would freeze over, and I got my iceskates out. The pond was now an ice skating rink. Fox would chase me as I went round & round the pond.

After about a year, I was riding my bike patroling the area around the pond. With Fox at my side, I raced downhill toward the pond. I tried to stop, but my back tire went into a skid. I went face first into murky, shadowy waters. When I went down, a minute later I felt something pull on my shirt, I grabbed it, not knowing what to think, when I hit the surface, I saw that it was Fox, pulling on my shirt as if he was trying to save me. He was to little to save me if I was really drowning, but it was the thought that counts, I owe him one.

Overhead A3.3,O9, p. 2

Another year passed. One day my mom got home from the store, and she bought me a rubber raft. It was just a cheap one, but it was mine. I blew it up with a tire pump. It was just the right size for me & Fox. Out of respect for Fox, I named it the USS Fox and christened it right in the pond.

On sunny days, I would take the raft out & lay in the sun with Fox on my legs. One day, when I was asleep in the raft, the wind blew pretty hard and blew my raft right into a bunch of sticksand rocks, the USS Fox was given a sadsalute, and then was no more.

Another year passed, and this would be our last year by the pond. I admired and respected that pond more than I ever did that year. But, at long last, all good things must come to an end, we moved to another town. Fox & I still visit the pond, but it'll never be like them 3 years when she was mine.

Overhead A3.3,O9, p. 3

Skimpy
Writing Assessment Criteria

- **Is organization suitable to the writer's purpose?**

Rarely	Sometimes	Frequently	Extensively

- **Does the writing exhibit a sense of personal expression?**

Rarely	Sometimes	Frequently	Extensively

- **Does the use of specific detail add to clarity?**

Rarely	Sometimes	Frequently	Extensively

- **Does the final draft exhibit an awareness of appropriate usage, mechanics and grammar?**

Rarely	Sometimes	Frequently	Extensively

Overhead A3.3,O10

Sample Holistic Writing Criteria

Strong Command of Narration	Has an effective opening and closing that ties the piece together. Relates to the topic and has a single focus. Well-developed, complete response that is organized and progresses logically. Writer takes compositional risks resulting in highly effective, vivid responses. Very few, if any, errors in usage. Variety of sentences and/or rhetorical modes demonstrates syntactic and verbal sophistication. Very few, if any, errors in sentence construction. Very few, if any, errors in mechanics.
Generally Strong Command of Narration	
Command of Narration	Generally has an opening and closing. Relates to the topic and usually has a single focus; some responses may drift from the focus. Development may be uneven with elaborated ideas interspersed with bare, unelaborated details. Some responses are organized with little, if any, difficulty moving from idea to idea; other responses may ramble somewhat with clusters of ideas that may be loosely connected, but an overall progression is apparent. Some errors in usage, no consistent pattern. May demonstrate a generally correct sense of syntax; avoids excessive monotony in syntax and/or rhetorical modes; may contain a few errors in sentence construction. May display some errors in mechanics but no consistent pattern.
Some Sense of Narration	
Little or No Evidence of Narration	May not have a opening and/or closing. May state a subject or a list of subjects or may have an underlying focus that must be inferred. Details are general, may be random, inappropriate, or barely apparent. Some lengthier papers are disorganized and difficult to follow; they may show no sense of planning. May have severe problems with usage including tense formation, subject-verb agreement, pronoun usage and agreement, or word choice. May contain an assortment of grammatically incorrect sentences; may be incoherent or unintelligible.

Overhead A3.3,O11

Sample Task Specific Writing Criteria

Prompt: Write about a place that is special to you.

5	Clearly states why the place is special and includes at least 10 details that describe it or explain why it is special or meaningful. The details appeal to at least four senses. Makes the place come alive for the reader. The writing is honest, appealing and from the heart. The piece is rich with adjectives.
4	
3	Indicates why the place is special or meaningful and includes at least 5 details that describe it or explain why the place is special. The details appeal to at least two senses. The place is clear to the reader, but not particularly alive. The writing is from the heart in some places and not in others. Adjectives are used, but are not particularly inspired.
2	
1	No special place is selected or there is little indication of why the place is special. Details, when present, are loosely connected. The paper is boring. Adjectives, when used, are mundane.

Overhead A3.3,O12

Speaking Performance Criteria to Critique

Score	Language	Delivery	Organization
A = 5	Correct grammar and pronunciation are used. Word choice is interesting and appropriate. Unfamiliar terms are defined in the context of the speech.	The voice demonstrated control with few distractions. The presentation holds the listener's attention. The volume and rate are at acceptable levels. Eye contact with the audience is maintained.	The message is organized. The speaker sticks to the topic. The main points are developed. It is easy to summarize the content of the speech.
B = 4 C = 3	Correct grammar and pronunciation are used. Word choice is adequate and understandable. Unfamiliar terms are not explained in the context of the speech. There is a heavy reliance on the listener's prior knowledge.	The voice is generally under control. The speaker can be heard and understood. The speaker generally maintains eye contact with the audience.	The organization is understandable. Main points may be underdeveloped. The speaker may shift unexpectedly from one point to another, but the message remains comprehensible. The speech can be summarized.
D = 2 F = 1	Errors in grammar and pronunciation occur. Word choice lacks clarity. The speaker puts the responsibility for understanding on the listener.	The student's voice is poor. The volume may be too low and the rate too fast. There may be frequent pauses. Nonverbal behaviors tend to interfere with the message.	Ideas are listed without logical sequence. The relationships between ideas are not clear. The student strays from the stated topic. It is difficult to summarize the speech.

Overhead A3.3,O13

Advantages and Disadvantages of Task Specific Criteria

Advantages of task-specific criteria:

- Quicker to train raters.
- High rater agreement right from the start.
- Therefore, often used in large-scale assessment.

Disadvantages of task-specific criteria:

- Have to develop new criteria for each task.
- Showing them to students ahead of time "gives away" the answer.
- Can't use them for judging the work in portfolios because the content of each student's portfolio can be very different.
- Right answers that are not on the list are frequently counted as wrong.
- Do not make the rater think.

Overhead A3.3,O14, p. 1

Advantages and Disadvantages of General Criteria

Advantages of generalized criteria:

- New criteria don't have to be developed for each task.
- Teaching students criteria right from the start improves performance.
- They help students generalize what quality looks like from one task to the next.
- They make students think.
- It is very beneficial to ask students to help develop them.

Disadvantages of generalized criteria:

- They take longer to learn (because the rater has to learn how to apply them to a variety of tasks).
- Rater agreement, at least at first, is lower.

Overhead A3.3,O14, p. 2

Advantages and Disadvantages of Holistic and Analytical Trait Criteria

Holistic–Advantages

- Quicker to learn
- Quicker to score
- Quicker rater agreement

Holistic–Disadvantages

- Don't communicate well. Two vastly different papers can get the same score for different reasons.

Analytical Trait–Advantages

- Provide better diagnostic information
- Provide better feedback to students
- More useful as a teaching tool

Analytical Trait–Disadvantages

- Takes longer to learn well
- Takes longer to score

Overhead A3.3,O15

Informal Writing Inventory

Purpose

"The Informal Writing Inventory is a diagnostic procedure…The results of analyzing samples are brought together into categories that simplify a judgment about the presence, degree, and, to a limited extent, the cause of writing disability…

Prompt:

"To motivate students to write, 14 illustrations are included…

Scoring:

"The identification of writing disability depends on two relationships:

1. The error index, which is the ratio of erroneous writing to well-formed writing; and
2. The communication index, which is percentage of total errors that disrupt communication."

Overhead A3.3,O16

Activity 3.4

Assessing Learning: The Student's Toolbox

Purposes:

1. To practice matching different assessment tasks to the student learning to be measured
2. To practice developing ("opening up") performance tasks

Uses:

This is an intermediate level activity that can be used in **Chapter 1** to illustrate the need to match assessment method to the learning targets being assessed, and in **Chapter 3** to assist teachers to design assessment tasks that match various learning targets.

Prerequisites might include (a) an activity on the rationale for alternative assessment (e.g., **Activities 1.1**—*Changing Assessment Practices...*, **1.6**—*Comparing Multiple-Choice and Alternative Assessment*, or **1.12**—*Assessment Principles*); (b) **Chapter 1** text or **Activity 1.7**—*Target Method Match*; and (c) **Chapter 2** text or an activity, such as **2.3**—*Ms. Toliver*, illustrating integrating assessmenty tasks into instruction.

Rationale:

As assessment initiatives seek to provide educators with the tools to gain insights into complex learning outcomes, discussions of assessment methods and task type can be overwhelming. This brief activity uses a toolbox metaphor, developed by Dr. David Clarke[1], to clearly illustrate different learning targets that can be addressed by slightly changing what began as relatively straightforward test items.

Materials:

- Overhead projector, screen, blank transparencies, chart paper and pens
- **Overheads A3.4,O1**—*Assessing Learning Purposes* and **A3.4,O2**—*Mathematical Tools*

---

[1] David Clarke, Faculty of Education, University of Melbourne, Victoria 3052, Australia

- **Handout A3.4,H1**—*Mathematics as a Set of Tools*

Time Required:

20-30 minutes

Facilitator Notes:

1. (5 minutes) This is a whole group presentation to set the stage for more detailed activities on expanding or creating performance tasks. Use **Overhead A3.4,O1**—*Assessing Learning Purposes* to highlight the purposes of this activity.

2. (5 minutes) Show the first section of **Overhead A3.4,O2**—*Mathematical Tools* and make the point that "opening up" our existing mathematics (or other) tasks so they can assess more complex student skills does not have to involve massive development; rather, it is possible to open-up tasks with some careful but simple changes.

   Some additional points to make: Traditional assessment items have heavily focused on answering the question of whether students have a particular mathematical tool in their "toolbox." But complex problem solving and other powerful mathematical thinking and investigating that we now expect as outcomes for learning demand more than "possessing the tools"—they demand understanding, using the tools in real world contexts, and knowing which tools to use in particular situations.

3. (5-15 minutes) Continue revealing each box on **Overhead A3.4,O2** while reviewing the examples that illustrate each type of question.

   After looking at each box, ask participants to be clear on which learning targets each successive "opening" addresses (to reinforce the notion of target-method match). For example, you might say: In this activity, the learning targets to be assessed are stated in the vocabulary of the tool box metaphor. State the learning targets measured by each task in the language of your local content standards.

   Possible answers: Tool possession=knowledge; tool understanding=knowledge; tool application=reasoning, problem solving; tool selection=reasoning, problem solving, math connections, communication in math.

4. Close with an invitation to experiment with opening-up the tasks currently in use and pass out **Handout 3.4,H1**—*Mathematics as a Set of Tools*.

Variation: Trying Out Opened-Up Tasks

Materials:

- Party favor streamers, crepe paper or string
- Lined or unlined paper

Time:

45-60 minutes

Facilitator's Notes:

1. (5 minutes) While presenting each box on **Overhead A3.4,O2**—*Mathematical Tools*, invite participants to actually do the tasks shown individually or with a partner. Prompt participants to discuss with their partner or with another pair the intellectual demands of each and provide additional examples of tasks aimed at answering the central question within each box.

2. (5 minutes) Following try out of the first task, *(Is the tool in the student's toolbox*?—"What is the average of 19, 21, 23, 25, and 27?") ask participants if it's possible to complete this task without understanding the concept of "average." (The answer is "yes.") Note that much of our traditional assessment is centered only on the "tool possession" questions.

3. (20 minutes) To carry out the second task, participants form into groups of 5 (the number is essential). Each group is given a streamer. This task ("Using only the streamer, find and display the average height of your group.") gets at the question: *Does the student understand the tool*? When groups are done, ask them to (a) describe what approach they took to the task, (b) whether each approach really leads the group to the "average" (rather than the median or mode), and (c) what more they might learn about the students' *understanding* of "average" following this task.

   (Note: Instead of the streamer activity, you might ask participants to do the other question in the box: "The average of five numbers is 23. What might the numbers be?" A useful extension of this task, underscoring emphasis on student understanding, is to ask participants if they can generate rules that could be used for similar average tasks. This results in a clear indication of whether the concept is understood. Some recent examples of generalizations: (a) "Add up all five numbers to get the sum. No matter what numbers you choose, as long as they add up to that sum, and you divide by 5, the average will be the same." (b) "Start with the number in the middle and go up and down equal increments on both sides." Ask the group to check the possible rule to see if it works.)

4. (5-10 minutes) The third task is "You are one member of a household of 5 people, whose average age is 23. What could be the ages of the other four people and who might they be?" It focuses on the question "*Can the student* *use* *the tool well in a real world context?*" Point out that while this is not an extremely "authentic" task, it is one that students respond to with great enthusiasm and energy—creating interesting family structures and then trying to communicate why certain ages are found among the members of the family. Ask participants how they might modify this task to match their own local context. Ask what they would know about student understanding of "average" from this task and what responses to the task say about students' ability to make mathematical connections between a real work situation and school math.

5. (10-15 minutes) The final sample task is one that demands students in small groups to independently *select* appropriate tools when presented with a situation that does not name the mathematical tool(s) that should be used. The task in the example is a Fermi Problem to which students must bring their experience, knowledge base and reasoning skills. Participants (and students who are given this type of task) are prompted with powerful closing questions: How good is your answer? Why?

6. You can extend this activity with an opportunity for participants to open-up tasks they have brought to the session or provide sample tasks that they can expand and critique for each other as critical friends.

Assessing Learning: The Student's Toolbox

Purposes:

1. To practice matching different assessment tasks to the student learning to be measured

2. To practice developing ("opening-up") performance tasks in assessment

Overhead A3.4,O1

Mathematical Tools

Tool possession

Our assessment can tell us whether a tool is in the student's toolbox:

"Find the average of 19, 21, 23, 25, and 27."

Overhead A3.4,O2, p. 1

485

Tool understanding

Our assessment can suggest whether the use of the tools is well understood as a procedure:

> *"The average of five numbers is 23, what might the numbers be?"*

or even as a concept:

> *"Using only the streamer provided, find and display the average height of your group."*

Tool application

Our assessment might even suggest whether the skills
can be applied in a real-world context:

> *"You are one member of a household of 5 people, whose average age is 23. What could be the ages of the other four people and who might they be?"*

Overhead A3.4,O2, p. 2

Tool selection

But a toolbox is of little value if the student never chooses to open it. Our assessment must tell us what mathematical tools the student chooses to use.

> *"It is Andrea's 13th birthday today. How many other children in the USA have their thirteenth birthday today?"*
>
> *How good is your answer? Why? What mathematical tools did you use to answer this question?*

David Clarke
Pacific Classroom Assessment Institute II,
Colonia, Yap, October 23, 1995.

Overhead A3.4,O2, p. 3

Mathematics as a Set of Tools

Tool possession

Our assessment can tell us whether a tool is in the student's toolbox:

> "Find the average of 19, 21, 23, 25, and 27."

Tool understanding

Our assessment can suggest whether the use of the tools is well understood as a procedure:

> "The average of five numbers is 23, what might the numbers be?"

or even as a concept:

> "Using only the streamer provided, find and display the average height of your group."

Tool application

Our assessment might even suggest whether the skills can be applied in a real-world context:

> "You are one member of a household of five people, whose average age is 23. What are the ages of the other four people and who might they be?"

Tool selection

But a toolbox is of little value if the student never chooses to open it.
Our assessment must tell us what mathematical tools the student
chooses to use.

> "It is Andrea's thirteenth birthday today. How many other children in the USA have their thirteenth birthday today?"
>
> How good is your answer? Why?
>
> What math tools did you use to answer this question?

David Clarke,
Pacific Classroom Assessment Institute II
Colonia, Yap, October 23, 1995.

Handout 3.4,H1

488

Activity 3.5

Performance Tasks & Criteria: A Mini-Development Activity

Purposes:

1. To understand that good assessment is also good instruction and vice versa
2. To experience the issues in using and designing performance assessments
3. To practice developing performance assessments

Uses:

This is an advanced level activity that can be used in **Chapter 3** as a summary of issues related to performance assessment design, and in **Chapter 2** to illustrate the integration of assessment and instruction.

Prerequisites might include: (a) an activity on the rationale for alternative assessment (e.g., **Activities 1.1**—*Changing Assessment Practices...*, **1.6**—*Comparing Multiple-Choice and Alternative Assessment*, or **1.12**—*Assessment Principles*); (b) **Chapter 1** text and **Activities 1.2**—*Clear Targets* and **1.7**—*Target-Method Match*; (c) **Chapter 2** text and **Activities 2.1**—*Sorting Student Work* or **2.3**—*Ms. Toliver* and **2.4**—*Is Less More?*, illustrating integrating assessment tasks and criteria into instruction; (d) one of the activities providing a gentle introduction to quality issues, such as **1.5**—*Clapping Hands* or **1.11**—*Assessment Principles*; and (e) the assessment quality activities in this chapter: **3.1**—*Performance Criteria, Keys to Success* and **3.3**—*Performance Tasks, Keys to Success* or **3.6**—*How to Critique an Assessment*.

Rationale:

Educators recommend experiential learning as a powerful tool for learning. It works well for adults too! This activity in developing performance tasks and performance criteria elicits myriad questions related to instruction and assessment.

Developing a performance assessment is a complex task in which a wide range of skills and knowledge are blended together. Initially, it can seem daunting to a teacher who wants to use performance assessment in his or her instruction. It is important to validate what accomplished teachers already do in their classes and to provide participants with the opportunity to see how a structured approach to developing alternative (performance) assessments can strengthen teaching and better link learning objectives to instructional activities and assessment. As one teacher put it, "I am beginning to understand that there is a blurring between what is good assessment and good instruction."

Materials:

- Overhead projector, blank transparencies and transparency pens, flip chart, and marking pens
- Miscellaneous commonly found things that could be used for a wide variety of activities, such as: paper, glue, scissors, tape, marking pens, rubber bands, plastic containers, string, a pile of leaves, sand, ... (You get the idea)
- **Overheads A3.5,O1**—*Performance Tasks and Criteria Purposes*, **A3.5,O2**—*How Tasks Differ*, **A3.5,O3**—*Overview of the Steps for Developing a Performance Assessment*, **A3.5,O4**—*Science and Math Performance Task Objectives*, **A3.5,O5**—*Sample Performance Task: How does grass get energy to grow?* **A3.5,O6**—*Discussion Questions*
- **Handouts A3.5,H1**—*How Tasks Differ*, **A3.5,H2**—*Criteria for Performance Tasks*, **A3.5,H3**—*Task Description Form*
- Sample performance criteria and scoring rubrics; suggestions are provided in the **Activity 3.5 Index** in **Appendix A**—*Sampler*

Time Required:

2-3 hours

Facilitator's Notes:

1. (2 minutes) Use **A3.5,O1**—*Performance Tasks and Criteria Purposes* to describe the purposes of the activity, as needed.
2. (10 minutes) Begin the activity by asking participants to do a word association game to help define *performance task*. In pairs, have them write down three words that they associate with the words *performance task*. Make a group list of these words and then ask participants to talk to their partners and come up with examples of performance tasks

that they use or have seen used in a class. Add these examples to the list and then help the group reach a shared definition of *performance task.*

> ***Performance Task:*** *Whatever activity, project, problem, or assignment that is given to students to do. For example, a particular math problem to solve, a lab to do, or a paper to write.*

Refer to the **Overhead A3.5,O2**—*How Tasks Differ* to talk briefly about how performance tasks may differ (also **Handout A3.5,H1**).

3. (5 minutes) The remainder of the steps in this activity revolve around having participants actually develop a performance task and a scoring system. Use **Overhead A3.5,O3**—*Overview of the Steps* to review with the group the steps in developing an alternative (performance) assessment.

Overview of the Steps for Developing an Assessment

- **Step 1:** Determine or clarify what you want to assess (see **Chapter 1**, *Keys to Quality Assessment,* **Key 1**—*Clear Targets).* What do you want to know about your students? As a reminder, participants can consult the National Standards, state documents, district objectives, or departmental guidelines when thinking about what to assess. Since we are discussing alternative assessment in this activity, participants should pick targets best assessed using this method. (See **Chapter 1**, **Key 3**—*Target Method Match* and **Activities 1.2**—*Clear Targets* and **1.7**—*Target-Method Match.*) Such targets tend to be complex, like skills, products, application of knowledge, and reasoning.

- **Step 2:** Develop or select a task (activity) that is well matched to what you want to assess. By this we mean that the task elicits the desired performance on the part of the student. For example, to assess critical thinking, the task must require students to think critically.

- **Step 3:** Identify the performance criteria. For example, if the purpose of the assessment is to assess critical thinking, there should be criteria for judging the quality of the critical thinking in student responses. What are the indicators and components of critical thinking? If the target to be assessed is conceptual understanding, what would students look like who really understand the concept?

- **Step 4:** Develop a scoring system. For this activity this means a rubric with several different performance levels for each performance criterion. Develop these or choose from the samples provided (see that suggestions in the **Activity 3.5 Index** in **Appendix B**—*Sampler*).

- **Step 5:** Decide on how feedback will be provided (e.g., grading, student self evaluation, verbal report, etc.)

- **Step 6:** Pilot test and revise

Although these are listed as distinct steps, it is important to note that some of these steps should be done in concert. For example as you are developing a performance task, you must also be thinking about the criteria you will use for assessment. In this way, identifying the criteria can inform the development of the task and vice versa.

For the purpose of this activity, participants will not be doing Steps 1 and 5. See other activities in the Toolkit98 to further develop these ideas.

4. (20-30 minutes) In groups of 3-6, participants will develop a mini performance task (see Step 2 above) based on a **given** learning target (Step 1). The task must be something that another group of teachers would be able to try out using the miscellaneous materials available and 20 minutes time. (Sometimes a quick [5 minute] scavenger hunt provides additional materials quickly.)

   You may want to choose a learning target that is specifically matched to the needs of the group or you may use the sample objectives provided (see **Overhead A3.5,O4**—*Science and Math Student Learning Targets*).

   To illustrate what participants must do, a sample performance task developed by teachers in another workshop, is shown on **Overhead A3.5,O5**—*Sample Performance Task: How Does Grass Get Energy to Grow?* The group can evaluate this task using **Handout A3.5,H2**—*Criteria for Performance Tasks* (introduced in **Activity 3.1**). This handout will be useful as the participants develop and evaluate their own performance assessments.

   Participants can use **Handout 3.5,H3**—*Task Description Form* to design their assessment.

5. (15-20 minutes) Exchange the activities developed by each group. The new group tries out the activity. This means actually following the instructions given to them and producing what is required (Step 6—*Pilot Testing*).

6. (20-30 minutes) Each small group can share with the whole group what the performance task was that they undertook, and explain what they did. This is a good chance for participants to ask questions of each other.

   Don't be surprised if a group interpreted the instructions differently than was intended. In an experiential way these conversations inform participants about what it is like to be a student and what information is important when doing an activity. This is the point of pilot testing—to ensure that tasks accomplish what we want.

   The key to the success of this activity is using the learning opportunities that arise and making teachers aware of broader issues through the examples that surface.

7. (10 minutes) Next, refine the performance criteria that were originally identified in the task design (Assessment Development Steps 2 and 3). If necessary, brainstorm some possible performance criteria and come to a shared definition of *performance criteria*. (e.g., science process skills, communication skills, knowledge of XXX, group process, etc.)

   Performance criteria: *a description of the characteristics that define how the performance on a task will be judged.*

   Give an explanation of how to develop a rubric with various criteria for quality and several (3-5) performance levels (see **Activity 2.1**—*Sorting Student Work*). Use the sample rubrics chosen using the **Activity 3.5**—*Index* in **Appendix A**—*Sampler* to discuss the different approaches that can be taken.

   Describing performance levels is not a trivial task. Having to formalize criteria into a scoring rubric will highlight the difficulties of distinguishing between performances. For example, you might notice that the highest performance levels are described in a positive way, whereas the performance levels described below it use more negative language (e.g., "plotted two of the graph coordinates wrongly and ....."). You can use this as a learning opportunity to discuss how a student might feel to have negatively worded feedback rather than a positively worded statement.

8. (15-20 minutes) Have the groups that did the performance task give their resulting work back to the group that designed the task. In their original small groups, participants can re-visit the performance criteria they had hoped could be assessed and, using the example response provided by the other group, see if the evidence of the criteria was present in the response (Step 6—*Pilot Testing*). The group should then determine if any changes to the criteria or to the performance task are necessary.

   (Note: The group will not actually give a score to the work produced when another group tried out their activity, however looking at this work is very useful in refining a scoring rubric. With actual work in front of them, participants will be able to see what information can truly be assessed and what can not. They can also talk about what is valued in the performance and perhaps what additional information would be helpful for their "students" to know in order to have the best opportunity to succeed on the task.)

9. (10-20 minutes) Many questions come up as a result of this activity, so it is helpful to have a whole group sharing of questions/issues. A group debriefing also provides a forum for clarifying issues and summarizing major ideas about developing alternative assessments. The following questions can be used to facilitate the conversation, or ask the participants to generate a list of their own questions.

   - What are some of the challenges of developing a performance task and/or performance criteria?
   - What are some lessons learned?

- It is time consuming to develop a scoring rubric that is specific to each performance task. What are some ways to maintain clear criteria without spending so much time for each task? (This can link to **Activity 3.3**—*Performance Criteria, Keys to Success* or **1.9**—*Going to School* and task-specific criteria versus general criteria. By comparing different sample rubrics, teachers can easily see how they differ and how a general rubric can be used for a variety of performance tasks.)
- Has this activity changed any of your thoughts about performance assessments and how you can integrate them with your curriculum?

Performance Tasks & Criteria Mini-Development Activity

Purposes:

1. To understand that good assessment is also good instruction and vice versa

2. To experience the issues involved in using and designing performance assessments

3. To practice developing performance assessments

Overhead A3.5,O1

How Tasks Differ

- One right answer vs. more than one
- Written, oral and/or visual instructions/activities
- Amount of choice on how to respond—written, oral and/or visual

- Format: short answer, essay, project, portfolio
- Specifying each skill that the student must use or allowing students to select whatever skills they think appropriate
- Individual and/or group work
- Timed or untimed
- Student choice on which task to complete
- What else?

Overhead A3.5,O2

Overview of the Steps for Developing a Performance Assessment

1. Determine or clarify the knowledge and skills you want to assess
2. Develop or select a task (activity) that is well matched to what you want to assess (e.g., to assess critical thinking, the task must require students to think critically)
3. Identify the performance criteria
4. Develop a scoring system
5. Decide how to provide feedback on performance
6. Pilot test and revise

Overhead A3.5,O3

497

Science and Math Performance Task Learning Targets

Science Learning Target to Assess

Develop an activity that requires students to demonstrate their understanding that science deals with processes of change, such as energy transfer, growth and aging, seasons, water cycles, etc.

Math Learning Target to Assess

Develop an activity that shows students' ability to apply a mathematical concept such as ratios, counting, sorting...to a problem that is relevant to their own lives.

Sample Performance Task

Title: How does grass get energy to grow?

Purpose: The learning targets addressed by this task are:

- students' group process skills
- students' understanding of different types of energy transfer
- students' ability to identify everyday examples of energy transfer
- student understanding that science is a process of change

Instructions: Make a collage, using both manmade and "natural" materials, to show at least three everyday examples of energy transfer. Prepare an explanation of your collage—each member of the group must participate in the explanation of how this is an example of how science involves processes of change

Products: One collage and a group explanation

Materials: Anything available on the table or that can be collected from nearby in 5 minutes or less

Criteria: You will be assessed on the following:

- Ability to work as a group
- Effective verbal communication skills
- Understanding of energy transfer by (a) applying knowledge to identify everyday examples, and (b) relating energy transfer to the broader topic of "science as a process of change"

Overhead A3.5,O5

Discussion Questions

- What are some of the challenges of developing a performance task and/or performance criteria?
- What are some ways to maintain clear criteria without spending so much time for each task?
- What are some lessons learned?
- Has this activity changed any of your thoughts about performance assessments and how you can integrate them with your curriculum?

Overhead 3.5,O6

How Tasks Differ

- One right answer vs. more than one
- Written, oral and/or visual instructions/activities
- Amount of choice on how to respond—written, oral and/or visual
- Format: short answer, essay, project, portfolio
- Specifying each skill that the student must use or allowing students to select whatever skills they think appropriate
- Individual and/or group work
- Timed or untimed
- Student choice on which task to complete
- What else?

Handout A3.5,H1

CRITERIA FOR PERFORMANCE TASKS
"Meatiness Scale"

Here is a list, adapted from one developed by Steve Leinwand and Grant Wiggins, of some important characteristics that we should look for in tasks. We will find it helpful to return often to this list, reviewing the quality of the curriculum and assessment activities we present to students.

Very "Meaty" ←——————————————→ Not "Meaty"
5 4 3 2 1

ESSENTIAL	• The task fits into the core of the curriculum. • It represents a "big idea."	vs.	TANGENTIAL
AUTHENTIC	• The task uses processes appropriate to the discipline. • Students value the outcome of the task.	vs.	CONTRIVED
RICH	• The task leads to other problems. • It raises other questions. • It has many possibilities.	vs.	SUPERFICIAL
ENGAGING	• The task is thought-provoking. • It fosters persistence.	vs.	UNINTERESTING
ACTIVE	• The student is the worker and decision maker. • Students interact with other students. • Students are constructing meaning and deepening understanding	vs.	PASSIVE
FEASIBLE	• The task can be done within school and homework time. • It is developmentally appropriate for students. • It is safe.	vs.	INFEASIBLE
EQUITABLE	• The task develops thinking in a variety of styles. • It contributes to positive attitudes.	vs.	INEQUITABLE
OPEN	• The task has more than one right answer. • It has multiple avenues of approach, making it accessible to all students.	vs.	CLOSED

Adapted from "Criteria for Performance Tasks," from *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*, edited by Jean Kerr Stenmark. Reston, VA: National Council of Teachers of Mathematics, 1991. Reprinted with permission. **Handout A3.5,H2**

Task Description Form

Title:

Purpose:

Instructions:

Product:

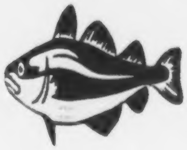
Materials:

Criteria:

Handout A3.5,H3

Activity 3.6

How to Critique an Assessment



Purposes:

1. To practice determining how well an assessment aligns with curriculum and instruction
2. To build skill using the **Chapter 3** guidelines for judging the quality of alternative assessment instruments: identifying content and skill coverage, the quality of performance tasks and criteria, equity and fairness, the consequences of assessment, and cost effectiveness

Uses:

This is an advanced level task. It can be used in **Chapter 3** to practice the skills needed to be a critical consumer of alternative assessments.

Prerequisites include: knowledge of the characteristics of a quality assessment (**Chapter 3** text, or **Activity 1.5**—*Clapping Hands*, **Activities 3.1**—*Performance Tasks, Keys to Success*, and **3.3**—*Performance Criteria, Keys to Success*); understanding the role of performance assessment in instruction, learning, and assessment (**Chapter 2** text or **Activities 3.2**—*Spectrum of Assessment Activity* and **2.1**—*Sorting Student* Work).

Rationale:

Using alternative assessment in the classroom does not automatically ensure good assessment. Students have a greater chance at success when learning goals, instructional design, and assessment methods complement each other. Even when instruction, assessment, and curriculum are aligned, it is possible to write assessment tasks that are flawed. The goal of this activity is to have participants discuss the quality of the Aquarium Problem (**Appendix A**—*Sampler*, **Sample A.15**) and whether it aligns to the curriculum content and the instruction modeled.

Materials:

- Overhead projectors, screen, overhead pens, chart paper, tape, and markers
- **Overheads A3.6,O1**—*How to Critique an Assessment Purposes*, **A3.6,O2**—*Critiquing a Performance Task*, **A3.6,O3**—*A Teacher Wants to Assess...*, **A3.6,O4**—*Critiquing the Aquarium Problem*, **A3.6,O5**—*Summary*
- **Handouts A3.6,H1**—*Alternative Assessment Evaluation Form*, **Appendix A**—*Sampler*, **Sample A.15**—*Aquarium Problem*

Time Required:

1-2 hours; optional extension requires an additional 30 minutes

Facilitator's Notes:

1. (5-10 minutes) Introduction

   a. Review purposes as needed (**A3.6,O1**—*How to Critique an Assessment Purposes*). You might note that the general procedure is to discuss the importance of quality assessment and how good assessment instruments align to the curriculum and the instruction modeled. For example, if students are taught a concept using problem solving, then the assessment should reflect the problem-solving skills practiced during instruction.

   This activity also gives participants practice at reviewing assessment instruments for quality and alignment. Teachers do not have time to develop all the assessment instruments they need, so they need to be critical consumers of those developed by others.

   Then say:

   > *We are going to critique an assessment and our job is to decide if this is a good assessment instrument.* (See **Overhead A3.6,O2**—*Critiquing a Performance Task*)

   b. Explain to the participants that a teacher has just finished a unit of study designed to build the following skills (learning targets): mathematical operations, problem-solving skills, communication skills, and data interpretation skills covered (see **Overhead A3.6,O3**—*A Teacher Wants to Assess...*).

   c. The teacher received the Aquarium Problem (**Appendix A**—*Sampler*, **Sample A.15**—*Aquarium Problem*) from the New Standards Project and he/she wants to determine if this would be a good assessment instrument to use with his/her students to measure the skills. Remind the participants that teachers do not necessarily need to develop every assessment instrument they use; they should beg, borrow, and steal assessments. Doing this requires critical review of the assessments of others to make sure the instrument is usable for their specific needs.

2. (5-10 minutes) Introduce the aquarium task (**Appendix A**—*Sampler,* **Sample A.15**—*Aquarium Problem.*

   a. Let participants know that the Aquarium Problem is from the New Standards Project from the University of Pittsburgh and is used with fourth grade students.

   b. Have participants look at the aquarium task—the problem, the scoring guide, and the five student responses. Participants should work the problem. You might ask if anyone has any questions about the task.

3. (5-30 minutes) Have participants review **Handout A3.6,H1**—*Alternative Assessment Evaluation Form* and point out the issues addressed. Ask the participants to read over the form and see if they have any questions about the criteria. It's best if participants have been introduced previously to notions of quality. (See prerequisites on page 1.)

4. (15-30 minutes) Critique

   a. You may use this as an individual activity, but it is recommended that participants are divided into five discussion groups. Each group could critique the assessment using all the criteria on the evaluation form, or each group could focus on a different section of criteria. Using **Overhead A3.6,O4**—*Critiquing the Aquarium Problem*, review their task with the participants:

   In your discussion groups:

   - Decide if this problem will elicit the skills the teacher wants to assess.
   - Using the alternative assessment form (**Handout A3.6,H1**), rate the quality of the assessment.
   - When the group has finished critiquing the assessment, list (on chart paper) the ratings agreed on for the assessment and the reasons for them. Add the things the group liked about the assessment and ways the assessment could be improved. Post the chart paper on the wall.
   - Decide on a spokesperson to report your results to the whole group.

506

b. As the groups work, walk around the room and listen to the conversations within each group.

5. (15-30 minutes) Discussion

a. After all groups are finished with the critiquing process, give each group time to report their findings.

b. Once the group report is complete, allow the other groups to ask questions or add additional comments to the presenting group's findings.

Be sure all the following points are covered:

- Did this assessment cover all the skills to be assessed? Did the task elicit the skills? Did the criteria assess them? What would you change?
- What are the possible sources of bias and distortion in this task?
- What could a teacher do with the results of this assessment?
- Would this instrument be worth the time spent by the teacher and the student?
- How could this task be modified to exhibit more authenticity?
- What type of curriculum coverage and instructional design was needed prior to this performance assessment?

6. (5 minutes) Summary

Show **Overhead A3.6,O5**—*Summary*. Ask participants if they have any more discussion or issues to address about critiquing an assessment instrument. Reiterate the major points of the session.

a. It is important to have experience at critiquing assessments prior to using them with students.

b. It is okay to change or modify instruments.

c. There are guidelines available to use for critiquing assessment instruments; however, particular circumstances might require emphasis on different things. Teacher groups could design their own critiquing instruments using the criteria they feel are most important.

d. Good alternative assessments cover the desired content and skills, have good performance tasks and criteria, are equitable and fair, lead to positive consequences, and are cost-effective.

Optional Extension

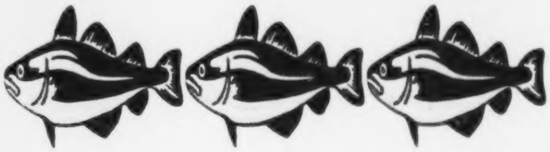
The **Fairness and Rater Bias** section on the alternative assessment evaluation form (**Handout A3.6, H1**) usually creates the most concern when discussing the Aquarium Problem. The issue is that the task context may not be equally familiar to all students. Many times teachers to not realize the need to discuss assessment tasks with others to determine if they are fair for all students. This may leave them feeling isolated and frustrated. The video mentioned in the reference box can be used to address these concerns.

🗁 Reference Box

The video *Effective Assessments—Making Use of Local Context* is available from WestEd, 730 Harrison Street, San Francisco, California 94107, (415) 565-3000.

Use video section ONE—*A Fish Story*. It shows how a group of Navajo teachers changed the aquarium problem to better suit the cultural background of their students while still assessing the same skills.

How to Critique an Assessment



Purposes:

1. To practice determining how well an assessment aligns with curriculum and instruction

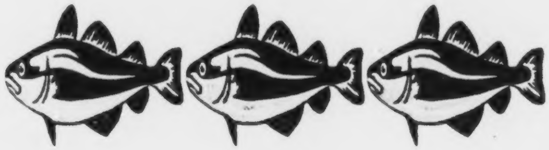
2. To build skills using guidelines for judging the quality of alternative assessment instruments:

   - content and skill coverage
   - the quality of performance tasks
   - the quality of performance criteria
   - equity and fairness
   - the consequences of assessment
   - cost effectiveness

Overhead A3.6,O1

509

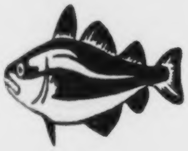
Critiquing a Performance Task



What are the strengths and weaknesses of this assessment instrument?

Overhead A3.6,O2

A teacher wants to assess the following knowledge and skills using a performance assessment



- Mathematical operations
  - ❖ adding, multiplying, dividing, subtracting

- Problem-solving skills
  - ❖ spatial relationships
  - ❖ reasoning

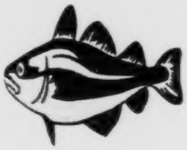
- Communication skills
  - ❖ letter writing
  - ❖ expression, clarity of explanations

- Data interpretation skills

Overhead A3.6,O3

511

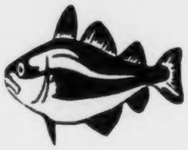
Critiquing the Aquarium Problem— A Sample Performance Assessment



- In groups, decide if this problem will elicit the desired skills
- Using the Alternative Assessment Evaluation Form, rate the assessment
- On chart paper, list ratings, things you liked, and recommendations for changes, if any
- Report out

Overhead A3.6,O4

How to Critique an Assessment Task—Summary



- It is important to have experience at critiquing assessments
- It is okay to change or modify assessments for classroom use
- There are guidelines available to use for critiquing assessment instruments

- Good alternative assessments:
  - ❖ Cover the desired content and skills
  - ❖ Have good performance tasks
  - ❖ Have good performance criteria
  - ❖ Are equitable and fair
  - ❖ Lead to positive consequences
  - ❖ Are cost effective

Overhead A3.6,O5

Alternative Assessment Evaluation Form

	Yes	Somewhat	No
1. Content/Skill Coverage and Correct Method	3	2	1

The assessment:

- Clearly states skills and content to be covered
- Correctly uses alternative assessment to measure these skills and content
- Avoids irrelevant and/or unimportant content
- Deals with enduring themes or significant knowledge
- Matches statements of coverage to task content and performance criteria

	Yes	Somewhat	No
2. Performance Criteria:	3	2	1

- Include everything of importance and omit irrelevant features of performance
- State criteria clearly and provide samples of student work to illustrate them
- Are stated generally, especially if the intent is use as an instructional tool
- Are analytical trait, especially if the intent is use as an instructional tool

	Yes	Somewhat	No
3. Performance Tasks:	3	2	1

General:

- Elicit the desired performances or work;
- Recreate an "authentic" context for performance
- Exemplify good instruction
- Are reviewed by others (students, peers, experts)

Sampling/Representativeness/Generalizability:

- Cover the content or skill area well; results can be generalized
- Sample performance in a way that is representative of what a student can do

Bias and Distortion:

- Avoid factors that might get in the way of student ability to demonstrate what they know and can do

Handout A3.6,H1, p.1

	Yes	Somewhat	No
4. Fairness and Rater Bias	3	2	1

Performance Tasks:

- Have content and context that are equally familiar, acceptable, and appropriate for students in all group
- Tap knowledge and skills all students have had adequate time to acquire in class
- Are as free as possible of cultural, ethnic, or gender stereotypes
- Are as free as possible of language barriers

Performance Criteria and Rater Training:

- Ensure that irrelevant features of performance do not influence how other, supposedly independent, features are judged
- Ensure that knowledge of the type of student does not influence judgments about performance quality
- Ensure that knowledge of individual students does not affect judgments about performance quality

5. Consequences	3	2	1

The assessment:

- Communicates appropriate messages
- Results in acceptable effects on students, teachers, and others
- Is worth the instructional time devoted to it; students learn something from doing the assessment and/or using the performance criteria
- Provides information relevant to the decisions being made
- Is perceived by students and teachers as valid

6. Cost and Efficiency	3	2	1

The assessment:

- Is cost efficient—the results are worth the investment
- Is practical/"do-able"

Activity 3.7

Chickens and Pigs: Language and Assessment

Purposes:

1. To increase awareness of how language and experience can influence assessment results
2. To gain expertise about equity issues and considerations when designing, using, and interpreting assessments

Uses:

This is an awareness level activity that can be used in **Chapter 1** to illustrate potential sources of bias and distortion, or in **Chapter 3** as an opener for a much deeper examination of the interaction of culture, language, and assessment.

Rationale:

This awareness activity emphasizes the critical role that language and culture play in effective and equitable assessments. It calls attention to a sometimes hidden dimension of equity in assessment—different meanings for the same term when interpreted in English and another language. Participants consider a common mathematics assessment task and generate their own ideas about language issues that arise when an assessment is used across languages and cultures. It is intended to prompt individual and small group reflection of this and similar issues.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A3.7,O1**—*Chickens and Pigs Purposes,* **A3.7,O2**—*Chickens and Pigs: Language and Assessment* (this can also be used as a handout), and **A3.7,O3**—*Discussion Questions*

Time Required:

10-20 minutes

Facilitator's Notes:

1. (2 minutes) Use **Overhead 3.7,O1**—*Chickens and Pigs Purposes* to introduce the purpose of the activity.
2. (5-10 minutes) Tell a story from your own experience (or use one of the stories below) regarding an equity issue that was hidden or difficult to recognize. Ask participants to consider a possible problem in the task that might prevent teachers from getting an accurate picture of students' learning, or which limits children's ability to portray their learning well. Here are three possible stories to get things started:

    a. Briefly introduce the "Chickens and Pigs" task (**A3.7,O2**—*Chickens and Pigs: Language and Assessment*) and identify it as a sample from early NCTM work designed to help teachers use a standards-based rubric to score student problem-solving responses.

    Show the task and call attention to the specific language problem that exists for some children and teachers. The following story from professional development work with teachers in Ulithi Atoll in the state of Yap, Federated States of Micronesia, can be used:

    Hoping to help teachers clarify what they value about student problem solving, and then build criteria from the key qualities of fine work identified in their small group discussions, the Chickens and Pigs task was given to all. Teachers began examining the task and some sample student responses. Discussing the task in their own language, puzzled faces turned to the facilitator and a senior member of the group asked if she realized that the word for legs in their language included both the legs and wings of a chicken! A lively discussion ensued about possible hidden factors that can mask student capabilities and understanding.

    b. A similar story comes from the city of Chicago. This time it's not language that blocks quality assessment, but the context and experience of the learners:

On a statewide assessment, children were asked to identify the number of legs on a chicken. Children from rural communities had no trouble providing the expected answer—two. City children, whose experience of chickens rested solely on those found in frozen packages in the grocery store, responded "four." When questioned, the children said that all the chickens they'd seen had four legs showing in the package.

c. Another example is a test item about birds sitting on a fence. Once again, context and experience blocked quality assessment, and in fact, the answer expected by the test developers was incorrect when examining the realities of rural life:

Four birds were sitting on a fence. A farmer threw a stone that hit one of the birds. How many birds were left on the fence?

(Every farm child knew that if there was a stone thrown toward the fence, no matter how many birds were hit, all would fly away and none would remain.)

3. (5-10 minutes) Ask participants to think of a time when they encountered a similar problem. Have participants examine the task (individually or in small groups) and discuss potential problems for students whose language and culture are not the same as that of the task developers. Use all or a few of the questions that follow (or create your own) to prompt individual reflection and then discussion (see **Overhead A3.7,O3**):

- What do you think are ways in which the task could pose problems for children of diverse cultures? Children whose first language is not English?
- As it is currently written, how much would you trust inferences about student achievement, knowledge, or skills from their responses to the task?
- How might this task be repaired to help gain a clearer picture of student ability? (**Extension**: Ask participants to underline parts of the task that could be changed to improve the task and then make changes that match your students' culture and environment.)
- What are some response options that do not require, or which minimize, written language requirements?

Chickens and Pigs: Language and Assessment

Purposes:

1. To increase awareness of how language and experience can influence assessment results

2. To gain expertise about equity issues and considerations when designing, using, and interpreting assessments

Overhead A3.7,O1

CHICKENS AND PIGS: LANGUAGE AND ASSESSMENT*

Problem: Ignacia and Cal went outside to feed the animals. They saw both chickens and pigs. Cal said: "I see 18 animals in all." Ignacia answered: "Yes, and all together they have 52 legs." How many chickens and how many pigs were in the yard?

---

* Adapted from NCTM

Overhead A3.7,O2

Discussion Questions

- What are ways the task could pose problems for children of diverse cultures? Children whose first language is not English?
- As it is currently written, how much would you trust inferences about student achievement, knowledge, or skills from their responses to the task?
- How might this task be changed to gain a clearer picture of student achievement?
- What are some response options that minimize written language?

Overhead A3.7,O3

Activity 3.8

Questions About Culture and Assessment

Purposes:

1. To expand awareness of cultural factors that affect students' ability to clearly demonstrate their learning during alternative assessments and our ability to interpret the quality of their work
2. To reflect on and discuss questions that are sometimes given little attention in assessment planning

Uses:

This is an awareness level activity that can be used in **Chapter 1** to illustrate potential sources of bias and distortion, or in **Chapter 3** as an opener for a much deeper examination of the interaction of culture, language, and assessment.

Rationale:

When alternative assessments are planned and implemented as part of systemic initiatives, there are profoundly important questions to ask and answer about culture. Examining the interaction of culture and assessment is essential in order to assure that *all* children have the best chance possible to create an accurate portrait of their knowledge and capabilities.

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens or chart paper and pens
- **Overhead A3.8,O1**—*Questions About Culture and Assessment Purposes*

- **Handout A3.8,H1**—*Critical Questions*

Time Required:

90 minutes

Facilitator's Notes:

1. (2 minutes) Use **Overhead 3.8,O1**—*Questions About Culture and Assessment Purposes*, as needed, to explain the purposes of the activity.

2. (10-20 minutes) One way to begin this activity is to open the session with the question—**What's the vision of the community for the future of its children?**

   In situations where parents and community members are invited to take part, this is a starting point that builds bridges between educators and families. Often the dreams that teachers, administrators, families, and communities have for their children are strikingly similar. Small role groups can brainstorm and share their answers, with the facilitator helping to emphasize the common dreams.

   (Note: In some cultures, separating the children from their community is inappropriate—they are part of the wholeness of the community. Thus, starting questions for vision building would include "What is the vision of the community for its future?" as well as "For the future of its children?")

3. (60-80 minutes) Following this consensus building, participants can be given handout **A3.8,H1**—*Critical Questions* and asked to select one or several questions for reflection and discussion. If your purpose is to create awareness of the multiple ways in which people answer these questions, cross-cultural groups might be effective. In all instances, individuals should be given the opportunity to self-select both the grouping preferred and the question(s). When participants work in "same culture" groups, you can use a jigsaw strategy that draws on the knowledge of each group and makes members responsible for sharing answers across cultures following the "same culture" discussions.

   Groups can consider all questions or divide up the questions depending on time available and whether or not this activity is part of an ongoing process for improvement. Groups may be brought back together for whole-group discussion if appropriate.

   (Note 1: Answers to the questions outlined in **Handout A3.8,H1** may differ even when participants and students share a common culture. The activity can then become a reminder of cultural factors to be consciously considered in alternative assessments. When participants bring separate cultural perspectives and experiences from those of their students, these critical questions need to be addressed in partnership with students and others whose knowledge of the culture(s) can be tapped. The language(s) of the discussions in this environment should be based on the needs of participants to communicate in the language in which they can best express their thoughts.)

(Note 2: When answers to critical questions are to be shared more widely and formally beyond the group, keep in mind that some participants may be free to speak for their culture, others may not. In several groups engaged in this process, members wished to consult with cultural leaders and respected elders to validate their preliminary responses.)

Discussion often raises points such as the following:

a. **The Role of Context:**

High context cultures (such as those found in the South Pacific and Micronesia) tend to use communication strategies in which most of the essential meaning is embedded in physical and non-verbal means. **Low context** cultures, by contrast, tend to use communications in which meaning is made explicit through detailed and sometimes elaborate verbal or written messages. What are the implications for assessment? Members of high context cultures interpret "over-elaboration" as an indication that the speaker believes they don't "get it." When the home culture values high context communication, the pressure for elaboration within school assessments presents a significant challenge to students. How do we build bridges for children across contexts so that they can be successful within the school culture without dishonoring their home culture?

b. **When Does Positive=Negative?**

Feedback about the quality of student work is greatly affected by the role of feedback in a culture. Positive feedback provided publicly to individuals or groups is inappropriate in some cultures. Overly positive feedback in some cultures (such as those in the South Pacific and Micronesia) is used to call attention to the weaknesses of the work. Awareness of the role of feedback, appropriate ways to give it, who may give it, and the power of positive or negative feedback is essential.

c. **What's the Role of Questioning?**

Who may ask questions differs widely across cultures. Some cultures (such as those in the South Pacific and Micronesia) consider student questions of the teacher as challenges to authority. Can distinctions be made between questioning to learn and questioning to challenge? How do efforts to promote critical thinking and inquiry impact culture and family relationships?

4. Facilitators and participants are encouraged to brainstorm additional critical questions. Some of the most fruitful discussions come from the questions raised by participants. This is an activity which can be very hard to close. Depending on your time limits, you can close by emphasizing that the handout is food for further thought and action.

Questions About Culture and Assessment

Purposes:

1. To expand awareness of cultural factors that affect a student's ability to clearly demonstrate learning during alternative assessments and our ability to interpret the quality of work

2. To reflect on and discuss questions that are sometimes given little attention in assessment planning

Overhead A3.8,O1

Critical Questions

This is a preliminary and incomplete list of questions about the complex relationship between assessment and culture. Many overlap. Some may be critical in some cultures, and relatively minor in others. Your additions, reactions, and suggestions are very welcome. What are your culture-assessment issues?

CULTURE AND ASSESSMENT

Assessment Question	**Related Questions and Issues**
What's worth knowing? Doing? Caring about? Assessing?	What knowledge is essential across cultures? What knowledge is essential within a culture? To what extent should/must "universal" knowledge take precedence?
	What's the vision of the community for the future of its children?
	How do we determine what's worth knowing in a culture? Who determines?
	What are the essential skills and habits of mind **across** cultures?
	What knowledge, skills, cultural practices, and values from within the students' culture must be assessed?
	Whose role is it to provide knowledge from within the culture? To what extent can schools take part without usurping authority in the culture?
	How do we develop quality assessments for deep cultural knowledge? What language or languages should be used? Who must be involved?
	To what extent is knowledge "owned" in the culture? What implications does this have for making the knowledge overt through instruction and assessment?

Assessment Question	Related Questions and Issues
What are the purposes for assessment?	What do teachers want assessments to do for them? For their students? For parents?
	Who are the key receivers of assessment information in the culture?
What are key qualities of fine work?	How are criteria determined? Who determines the criteria used for assessing student work?
	What does quality look like in the culture?
	Where complex cultural practices and knowledge are involved, how can teachers from another culture determine quality? Must assessment be carried out only from within?
How can students be prepared to display vivid images of their learning?	What are powerful metaphors and parables of learning and assessment in the culture?
	Are visual images useful? Is the culture one that values oral expression?
	What are the communication patterns and preferences? To what extent should assessment tasks honor those patterns? To what extent and at what level(s) should assessment tasks require students to stretch themselves beyond their comfort?
How will we know they're learning?	What are culturally appropriate ways for children to display their learning?
	What language will assure students opportunities to truly display their learning? What assessment purposes require the use of students' home language? A second language?
	To what extent should assessments require students to be multicultural?
	Is self-reflection a universal habit of mind? Is the revealing of self-reflection to others appropriate? Or is it more related to personality?

Assessment Question	**Related Questions and Issues**
	What are the key differences and similarities between the culture of the home and community, and the school culture?
What's a quality assessment task?	Should tasks be "culture free"? How much context is essential? Should we begin with local context and then design tasks that require knowledge of other cultures?
	What kinds of tasks are common in the culture?
	Is the culture one that emphasizes high-context communication? What are implications for assessment? For gauging the quality of student work?
	Who can work together in a group? Who cannot be in the same group?
	Will grouping by gender help or hinder?
Do you see what I see?	Are perceptions of quality the same across cultures?
	What are indicators of quality in the culture?
How should students be given feedback about the quality of their work?	What are appropriate forms of feedback in the culture? Is individual, public feedback viewed as positive? Very negative? Is non-public individual feedback effective? Acceptable?
	Is group feedback preferred? Are there strategies for providing individual feedback for work done in groups?
	Is verbal feedback effective? Written feedback?
	Who can give feedback? Who cannot? Is peer assessment acceptable?
	Is feedback in the culture focused on the strengths of the work? Is it focused on weaknesses? Do children from a culture that provides negative feedback respond to positive feedback?

Activity 3.9

Tagalog Math Problem

Purpose:

To experience the effect of one potential source of bias in testing—not knowing the language in which the assessment occurs.

Uses:

This is an awareness level activity that can be used in **Chapter 1** to illustrate potential sources of bias and distortion, or in **Chapter 3** as an opener for a much deeper examination of the interaction of culture, language, and assessment.

Rationale:

Bias in assessment occurs when a skill not being assessed interferes with the student's ability to show what he or she can do on the skill actually being assessed. In this case the math skill being assessed is averaging. But, the typical American student would not be able to demonstrate his or her ability to average because he or she would not know what the question asks. Thus, not knowing the language (a skill not being assessed in this case) inhibits the student's ability to show what he or she can do in math (the skill actually being assessed).

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A3.9O1**—*Tagalog Purposes*; **A3.9,O2**—*Tuusin…*; **A3.9,O3**—*Add…; A3.9,O4—Discussion Questions*

Time Required:

20-30 minutes

Facilitator's Notes:

1. (2 minutes) Use **Overhead A3.9,O1**—*Tagalog Purposes*, as needed to explain the purpose of the activity.

529

2. (5 minutes) Show the task in Tagalog (**Overhead A3.9,O2**) and ask participants to solve the math problem. If they ask for a translation tell them, "I'm sorry, I'm not supposed to give you any help on the test. Just do the best you can."

   After a minute or two ask participants what their solution is and why they approached the problem the way they did.

3. (5 minutes) Show the English translation (**Overhead A3.9,O3**) and ask participants to compare what they guessed they were supposed to do to what they were actually supposed to do.

4. (10-20 minutes) Discuss (use **Overhead A3.9,O4**):

   a. Was it fair to assess math in this way?

   b. Has a similar thing ever happened to themselves or a student they know? Share examples.

   c. What other things might bias the results of an assessment? (You might need to remind participants that bias in assessment occurs when something that you *are not* trying to assess interferes with something you *are* trying to assess. Such as, too high a reading level on a math test; too much writing on a reading test; ability to play a role when you're really assessing critical thinking skills; etc.

   d. Under what assessment circumstances might language be a biasing factor, and under what assessment circumstances would language not be a biasing factor? (Hint: If ability to use English is the skill being assessed, then language is not a biasing factor. If you are not assessing ability to use English, then language might be a biasing factor. The essential part is *to be really clear on what it is you are trying to assess.* If you aren't really clear on what skills are being assessed, it is impossible to tell what other skills required by the task might bias the results.)

Tagalog Math Problem

Purpose:

To experience the effect of one potential source of bias in assessment—not knowing the language in which the assessment occurs.

Overhead A3.9,O1

Tuusin ang suma ng lahat ng numero sa ibaba at kalkulahin ang promedyo.

14

18

27

25

33

22

16

13

Overhead A3.9,O2

Add up all the numbers below and calculate the average.

14

18

27

25

33

22

16

13

Overhead A3.9,O3

Discussion Questions

1. Was it fair to assess math in this way? Why or why not?
2. Has a similar thing ever happened to you or someone you know?
3. What other things can bias the results of an assessment?
4. When are reading and writing in English a biasing factor in an assessment, and when are they not?

Chapter 4

Grading and Reporting—A Closer Look

What's in This Chapter?

"How do I incorporate rubric scores into grades?" "How do I define standards for grades that are fair to all students?" "What do grades mean?" Questions such as these are common right now. Teachers at all grade levels are currently feeling a certain amount of discomfort about the manner in which (either by choice or by district policy) they have assigned grades. Many teachers have told us that the method by which they are required to grade (high school teachers, for example, frequently having to have, say, 15 marks during a grading period) prevents other changes from occurring in assessment and instruction. There is a sense that grading practices must somehow change, but no one is quite sure what must change or how—or even which questions about grading are the best to ask.

No one currently has *the* answer. However, in this chapter, we provide current best thinking about the issues that have to be addressed and information about how others have addressed these issues in redesigning grading and reporting systems.

Chapter Goals

1. Explore issues with respect to grading and reporting
2. Reinforce the importance of aligning grading and reporting strategies with valued learning targets and instructional strategies
3. Increase knowledge of strategies for incorporating alternative assessment data in the overall determination of student grades
4. Reflect on options for broadening communication with parents, students, other teachers, and the community about the quality of student work

Chapter Content

A. Readings

Introduction..........Page 4

Grading: Issues and Options..........Page 6

Describes a variety of approaches for incorporating alternative assessments into a grading scheme, and explores issues surrounding grading.

Reporting: Communicating About Student Learning..........Page 17

Examines assessment as a communication system that engages students, teachers, parents, and others in ongoing conversations about the quality of student work in order to lead to improvement. Different report card formats are described.

B. Activities

Activity 4.1 *Weighty Questions*

This advanced level activity helps to illustrate the importance of developing sound grading practices that reflect valued student learning targets. Time: 90 minutes to 2 hours; extension, 60 minutes

Activity 4.2 *Putting Grading and Reporting Questions in Perspective*

This intermediate level activity provides an opportunity for teachers to express their current questions and concerns about grading and relate them to three levels of concerns about grading proposed in a paper by Alfie Kohn in **Appendix C**. Time: 75 to 90 minutes

Activity 4.3 *Grading Jigsaw*

This intermediate level activity raises and promotes discussion about issues surrounding grading. It uses papers on grading and reporting in **Appendix C**. Time: 1 ½ to 2 ¼ hours

Activity 4.4 *Won't Some Things Ever Change?*

In this intermediate level activity, participants compare and critically examine the messages sent by report cards from 1916, 1943, 1965, and 1994. Time: 30-40 minutes.

Activity 4.5 *Report Card Formats*

This advanced level activity provides examples of innovative report card formats and promotes discussion of their relative advantages and disadvantages. Time 40-75 minutes

Activity 4.6 *How To Convert Rubric Scores to Grades*

In this advanced level activity, participants discuss the advantages and disadvantages of four procedures for converting rubric scores to grades. This case study focuses on writing. Time 60-75 minutes

Activity 4.7 *Case of the Killer Question*

This advanced level activity presents the real-life grading dilemmas faced by an alternative high school. The killer question is: "How do we capture and report the learning of our students without compromising the integration, authenticity, student autonomy and self-directed learning that's at the heart of our program?" Time: 1 ½ to 2 ¼ hours

Activity 4.8 *Guidelines for Grading*

This advanced level activity presents eight guidelines for grading taken from current research. Participants consider each guideline in light of the goals they have for grading. Time 1 ½ to 2 hours.

Activity 4.9 *Grading Scenarios*

This advanced level activity presents real-life grading dilemmas for discussion. Time: 90 minutes

Grading and Reporting–A Closer Look Readings

What Are the Issues in Grading and Reporting?

> *Grading is not assessment; it is one—fairly simplistic—form of coding assessment information.... There are many other ways of representing and exchanging assessment information. Assessment should not be identified exclusively with a single method of representing assessment information, and the process of assessing learning should certainly not be confused with the activity of coding the results of our assessment. When this confusion occurs, it merely invites students to confuse learning achievement with achieving a good score. Compound this with the lesson that the way to a good score is to anticipate the desired response, and we are well on the road to instilling a view of education that many of us lament in our students, perhaps not realizing how much our practices have encouraged them to adopt it.*
>
> (David Clarke, 1997, pp. 20-21)

Grading

Today, the things society values as important learning targets for students have changed. Developing and assessing "big outcomes" (problem-solving, reasoning, communication, critical thinking) are considered at least as important as transmitting and assessing factual knowledge. Numerous efforts are underway across the country to foster instruction that helps students develop these complex skills. Assessment of skills tends to be more open-ended and performance-based, requiring more subjective judgment in scoring than do multiple-choice, factual recall tests. Some teachers

🗀 **Reference Box**

David Clarke, *Constructive Assessment in Mathematics*, 1997, Berkeley, CA: Key Curriculum Press.

express confusion about how to fit these new methods into traditional percentage-based grading systems. Teachers who work within the constraints of having to give a single grade for six weeks, nine weeks, or a semester need suggestions for how to incorporate information from alternative assessments.

This chapter presents ideas for doing this. Frequently, however, in order to decide on the best course of action, grading has to be put into the larger context that considers both why grades are required and why teachers assess at all. What are grades supposed to do for us? Motivate students? Communicate about achievement? Provide rewards and punishments? Just what? Is there any research to show that grades are an effective means for doing these things?

Why *do* teachers assess? Just to report on student progress by grading? Or are there other reasons, like evaluating one's own instruction, planning instruction, and empowering students through self-assessment? Are there better ways to motivate students through assessment than promising "A's" and threatening "F's?" How do achievement targets for students, instructional philosophy, classroom assessments and grading practices line up? Until there are answers to these and similar questions, satisfying solutions to grading issues can be elusive. So, this chapter also presents these issues in light of current research.

A paper that does a nice job of setting the context and providing the big picture for grading is *Grading: The Issue Is Not How But Why* (article #1 in **Appendix C**—*Articles*.) Kohn proposes three levels of concern about grading. Level I concerns deal primarily with the "how to" questions such as "How do I combine numbers to form grades?" Level II concerns ask, "How do I use the information from alternative forms of assessment to make my grades more sound?" and "Are there other uses for assessment besides grading?" At this level teachers are not questioning the need to grade or assess, but recognize that grades need to be based on sounder evidence. Level III concerns focus on questions of, "Why do I grade and assess at all?" At this level teachers begin to question the reasons for grading and assessing and wonder if there might be a better way to motivate students and communicate about their achievement.

This chapter focuses on Level II and III questions.

Reporting

Reporting issues go hand in glove with grading issues. If the goal of assessment is to communicate about achievement status and progress with students and their parents, what is the best way to do it? Many schools use alternative methods for reporting student progress besides a single indicator (such as A, B, C, D, and F letter grades). They might use such things as skills checklists, verbal descriptions of performance, or developmental continuums to track progress.

But if report cards do use grades to communicate effectively to parents and students, the basis for their assignment should be clearly stated. It is not enough to say that 90 percent or higher is an A if the "90 percent of what" question isn't answered. This chapter suggests that we clearly define "the what" so that grades communicate effectively and accurately. Since our focus is on the classroom in this chapter, we focus on exchanging information between students and teachers, teachers and parents, and students and parents. This is not a *complete* treatment of all the reporting necessary for the schools. Educators live in a society where public schools are accountable to the tax paying public, just as businesses are accountable to their stockholders. Because of this, reporting student achievement also is driven by accountability needs. So, the emphasis in this chapter—reporting at the classroom level—also needs to be placed in the context of broader reporting needs. These broader needs are not discussed here; they will be the topic of future supplements to *Toolkit98*.

Grading: Issues and Options

According to Rick Stiggins, assigning a quality grade involves several steps.

Step 1: Spelling out the big achievement picture. Identifying clear targets—what do we want students to know and be able to do at the end of the grading period?

Step 2: Turning this big picture into an assessment plan in which learning targets are matched to assessment methods. What evidence will be gathered to show the extent to which student competencies have been attained?

Step 3: Moving from plans to assessments. The plan needs to be turned into actual assessments.

Step 4: Moving from assessments to information. The assessments must be given, scored, and results recorded.

Step 5: Summarizing the resulting information. How will different kinds of evidence be weighted and combined?

Step 6: Converting composite achievement scores to a grade. How will the composite information be converted into a grade?

Reference Box

For more information see Rick Stiggins, ***Student-Centered Classroom Assessment***, Columbus, Ohio: Merrill, 1997, Chapter 15.

Steps 1 and 2 were also discussed in **Chapter 1** (clear targets and target-method match). Step 3 was addressed in **Chapters 2** and **3** (designing assessments to fulfill different purposes). Here, we'll expand on Steps 1 and 2 in the context of grading and, particularly, grapple with Steps 5 and 6.

Steps One (Clear Targets), Two (Target-Method Match) and Five (Combining Information)

The first decision—identification of student competencies to be targeted—is the key to the process. If the competencies to be targeted in the class are not clear, the rest of the process will be fuzzy. If student goals are clearly stated, not only does instruction become much more focused, choices for assessment strategies and the grading process become much more logical.

Example 1. For example, sample student learning targets for a science class, and their associated assessment methods and weighting are shown in **Figure 4.1**. In this example, conceptual understanding will be assessed using paper-and-pencil and interactive computer tests (40 percent of the grade). Alternative assessments such as lab or performance tests, projects, papers, and journals will be used to assess science process skills, application of knowledge in creative ways, problem solving, and responsibility (60 percent of the grade).

Figure 4.1
Sample Unit Assessment Plan

Student Learning Targets	**Assessment Tasks—target-method match**	**Weighting in Final Grade**
Conceptual understanding of science content	Paper-and-pencil test; interactive computer test	40%
Science process skills	Performance tests	20%
Creative applications of knowledge	Paper-design a new solution based on knowledge	15%
Research and problem-solving skills	Extended project	15%
Responsible for own learning	Journal or portfolio, project	10%

This chart addresses Steps 1, 2 and 5 above: what student competencies will be developed, what assessment methods will be used, and how the assessment information will be weighted. Notice

541

that the assessment process becomes more logical when it is organized around student competencies and outcomes. And, notice that this becomes a useful communication tool for students and parents, clarifying the purposes of instruction. (But, the teacher may want to add some descriptions and definitions so every one knows specifically *what* is to be understood, and *which* skills will be developed.)

Example 2. Developers of the Hawaii Algebra Learning Project (**Sample A.31**—*Hawaii Algebra Grading Process* in **Appendix A**—*Sampler*) want to develop the student competencies shown in **Figure 4.2**. They also believed that it is important for students to work hard every day and not just study for tests. The following quote from *Assessment Standards for School Mathematics* articulates the latter concern:

> *Tests are just small tasks...They are used at a point in time, not as a reflection of growth over time. When we grade tests, we are only looking at first drafts because students have not had the opportunity to revise their work due to the time limitations under which the test is administered. The time spent taking a test is a very small part of the time spent using mathematics, so why should the test count a lot?* (p. 132)

Reference Box

National Council of Teachers of Mathematics Assessment Standards Working Groups, 1993. *Assessment Standards for School Mathematics,* (703) 620-9840,

Figure 4.2
Chapter Assessment Plan

Student Learning Targets	Assessment Tasks—Target-Method Match	No. Entries	Weight
Math problem solving	Student discussions and group presentations, open-ended problems	11	42%
Communication of mathematical thinking	Journal prompts, open-ended problems, discussions	11	42%
Conceptual understanding of math content	Quizzes, tests, open-ended problems	2	8%
Application of knowledge	Labs, open-ended problems	2	8%
	Total	26	100%

(Please notice that the same assessment tasks can be used to provide information on more than one learning target. For example, a rubric for "ability to communicate mathematical thinking could be used with the journal, open-ended problems, or group discussions.)

To address the learning targets of problem solving and communication of mathematical thinking, daily work in the Hawaii project consists of collaborative problem-solving and journal-writing tasks that are weighted more heavily than tests. For example, for a given time period, there might be 26 gradebook entries scored from 0 to 4 (see **Figure 4.2**). A score of 4 indicates that the student's work meets all the qualifications for excellent work (including excellent content understanding), comparable to an A. Students receive approximately 11 scores on their collaborative group's class presentations on assigned algebra problems. They receive 11 more scores representing the quality of their mathematical thinking in response to journal prompts such as "How is solving an equation like/unlike solving an inequality?" Finally, to assess conceptual understanding and application of knowledge, students receive one quiz score, one chapter test score, and two lab scores.

Hawaii project authors claim that emphasis on daily problem solving results in decreasing fear and anxiety about tests and increasing intrinsic motivation to understand important concepts and to be involved in class discussions. Additionally, students do not memorize for a test and then promptly forget the content after the test. Rather, they discuss and apply content information daily. Assessment is integrated into the instructional process.

Quality Issue Box

In order to make the Hawaii scheme work, and have credibility with parents, the teacher must have some way of consistently judging the quality of the student presentations, journal entries, and labs so that accurate scores are assigned. What credibility would such a system have if judgments varied between teachers or differed with the same teacher on different occasions? (The grade is only as good as the assessments on which it is based.) This is where performance criteria come in. (Sample A.31—*Hawaii Algebra Grading Process* has rubrics for scoring journal prompts and presentations.) Teachers have to practice using these scoring guides so that judgments are consistent across students, days, and assignments.

Also to make this work, teachers need enough *samples* of performances to make sure that the grade really represents true achievement. Hawaii teachers gather 11 samples of collaborative problem solving and 11 writing samples. We can all agree that one example of a student's ability to solve math problems is not enough to determine for sure how good the student is at solving problems. Is 11 enough? What would have to be true of those 11 samples for you to feel comfortable that the scores in the grade book really represent the student's ability to collaboratively solve problems? (Hint: coverage of problem solving strategies and content.)

BEST COPY AVAILABLE

Example 3. A high school English teacher, Margie Krest, uses portfolios to assign grades. Her grading scheme varies somewhat from class to class depending on the developmental level of the students and the skills on which they are working. **Figure 4.3** shows one example. The teacher wants her grading scheme to reinforce her student learning targets. For example, although students have to complete all assignments, not every piece of writing is "graded"; this encourages students to experiment with different genres and writing conventions (e.g., dialogue). If she graded every piece of writing, students would be less likely to experiment. Further, she rewards students for experimentation by directly incorporating risk-taking into her grading scheme. She also wants to reward revision, so students get better grades if there is evidence that they have thoughtfully revised their work.

Related Toolkit98 Chapters and Activities.

Chapter 1 has more detail on developing clear targets for students, choosing appropriate assessment methods for each target, and sources of bias and distortion in assessments (such as sampling). Related activities are: **Activity 1.2**—*Clear Targets*, and **Activity 1.7**—*Target-Method Match.*

Activity 4.1—*Weighty Questions* provides *Toolkit98* users an opportunity to develop an assessment plan similar to those in **Figures 4.1** and **4.2.**

Figure 4.3
Grading in a High School English Classroom

Student Learning Targets	**Assessment Tasks—Target-Method Match**	**Weight**
Writing fluency, revision, risk taking	Portfolio of sample writing containing all drafts, revisions, prewriting material, suggestions from classmates, and final drafts	50%
Quality of writing	Three writing samples selected by the student from the portfolio	50%
	Total	100%

Step Six (Converting Scores to a Grade)

Figures 4.1-4.3, above, show how targeted competencies, assessment methods, and weighting the relative importance of competencies can be matched-up. The last step in determining report card grades is the grade assignment process—how the final grades will be determined based on the information collected. The methods described below are taken from teacher practice and the literature on grading.

Fixed Percent Scale Method. In this approach, scores from each grading component (unit tests, quizzes, performance assessments, observations, homework) are expressed as a percentage. At the end of a grading period, these component grades are weighted and averaged (step three) and assigned a letter grade based on a fixed scale (e.g., 93 to100% = A). This is the approach currently used by many teachers.

> 🕮**Reference Box:**
>
> David Frisbie and K. Waltman, *Developing a Personal Grading Plan*, 1992, in *Instructional Topics in Educational Measurement.* Washington, DC: National Council on Measurement in Education, (NCME) (202) 223-9318.
>
> J. Ory and K. Ryan, *Tips for Improving Testing and Grading*, 1993, Newbury Park, CA: Sage Publications, Inc., (805) 499-0721.

In a classroom where the teacher is teaching to complex instructional learning targets and assessing these "big outcomes" with alternative methods, this traditional approach can result in several problems.

1. Knowing that 93 percent *always* means an A may lead to restricting what the test covers.

    > *If the teacher believes there should be some A grades, a 20 point test must be easy enough so that some students will score 19 or higher; otherwise there will be no A grades. This circumstance creates two major problems for the teacher as assessment developer. First, it requires that assessment tasks be chosen more for their anticipated easiness than for their content representativeness. As a result, there may be an over-representation of easy concepts and ideas, an overemphasis of facts and knowledge, and an under representation of tasks that require higher-order thinking skills.* (Frisbie & Waltman, 1992, p. 40)

    A student score of 80 percent might represent different performances on two different tests. It may be that one test assesses higher-order thinking and the 80 percent actually represents a better performance than 80 percent of a short-answer knowledge test.

2. Secondly, scoring guides (rubrics) for alternative assessments are intended to *describe* the level of a student's performance on a task. Converting the descriptive score to a percent score makes no sense. If five stars are the maximum possible and a restaurant receives a

BEST COPY AVAILABLE

three star rating, it would make no sense to convert the 3 to a 60 percent, and judge the restaurant a failure. Three stars may actually mean "OK" or even "good."

For example, consider **Sample A.16**—*Discovering the Problem of Solid Waste* in **Appendix A**—*Sampler*. It is scored on a variety of dimensions using a point system with the total points possible 11. The points are assigned based on how well students' responses match the desired responses—75 percent might actually indicate good (B) performance.

Also consider **Sample A.41**—*Writing in Chemistry* (in **Appendix A**—*Sampler*) which uses an analytic rubric to score responses to a chemistry writing task. Each essay is scored from 0 to 5 on six dimensions (general impression of content quality, use of prior knowledge, number of chemistry principles or concepts used, quality of argumentation, number of specific references to a source text, and accuracy of knowledge). The profile of scores communicates to students their strengths and weaknesses (e.g., strong on "argumentation," weak on "number of principles"). Deriving a percent score would only confuse the message to students. A student who received all 3's would fail the task with a 60 percent, using a percent scoring system, whereas 3's may represent minimally acceptable performance (C work) in the teacher's judgment.

For the above reasons, a direct conversion of rubric scores to percentages seems to be incompatible with alternative assessments. Other options are described below.

Total Point Method. Some teachers have students earn points throughout a grading period, then assign grades based on the point total at the end of the period. For example, the teacher may determine that for a particular grading period points will be accumulated as shown in **Figure 4.4**. Other examples: **Sample A.16**—*Discovering the Problem of Solid Waste* shows a performance test scored with a point system which *might* represent 11 of the 40 points for conceptual understanding. **Sample A.36**—*Assessment of Laboratory Skills in High School Science* is an example of how 35 points on a lab test *might* be accumulated to assess science process skills. Finally, to assess communication skills, a sixth-grade journal might be checked for completeness of entries each week with a maximum of 10 points per week.

Figure 4.4
Sample Point Assignments for a Grading Period

Student Learning Targets	Assessment Tasks—Target-Method Match	Total Points Possible
Conceptual understanding of science content	Paper-and-pencil tests Interactive computer test	40 50
Science process skills	One lab One demonstration summary	35 15
Creative applications of knowledge	Two tasks at 10 points	20
Plan/implement/share independent learning	One project, two creative application tasks, lab	50
Ability to communicate science understanding	Journal, lab demonstration summary, creative applications tasks, project	60
	Total	270

Point values for the assessment tasks and the proposed cutoffs for grade assignments should be determined prior to starting the grading period. For example, in **Figure 4.5** students might be told that 235 or higher would earn an A. The decision about where to draw the line for an A, B, C, D, or F depends on the teacher's standards and expectations about what students should accomplish, not necessarily on "90%=A."

Figure 4.5 shows how totaling points might be a more equitable way to assign grades than deriving percents and then averaging.

Figure 4.5
Totaling Points

Assignment #	Points Received	Converted to Percent
1	9 out of 11	80%
2	45 out of 50	90%
3	2 out of 5	40%
4	14 out of 20	70%
5	40 out of 50	90%
	110 out of 136	AVG.
	81% = C	74% = D

The point system approach may fit well in classrooms where teachers use rubrics. For example, students may be asked to write up science demonstrations several times per grading period. Each write-up may be worth 20 points (5 points each for hypothesis, method, results, and conclusion). A science project may be worth 100 points with points awarded for a variety of criteria. Another point system might be used for demonstrations of effective cooperative group skills.

This approach may also work well for teachers who not only want to grade work for quality, but want to provide incentives for students to complete their work. Here, the total points awarded over a grading period could be a combination of points for completing the work that is designed to be instrumental in developing a particular competency and points awarded for the level of performance on unit and final assessments.

For example, one middle school science teacher has her students keep a "science points score card," which is given out at the beginning of each class period, with opportunities during the class to earn points for completed work. Over a grading period, students have the opportunity to earn a large number of points (e.g., 600) both for completing work and for graded assignments or tasks. This approach may be particularly effective in the middle grades to decrease "test anxiety" and fears of failure and to keep students involved and feeling successful. Accumulating points for completing work, when the work has major instructional value, increases the likelihood that students will be successful and, thus, stay involved in the learning.

But, there's also a downside to totaling points as a way to assign grades. Assignments worth more points are weighted more than assignments with fewer points. That's great if this, in fact, is what's desired. But what about rubric scores which might total 5 points or fewer even for great work? In this case, converting all work to grades or percentages first and then averaging the grades or percentages might be a better reflection of student achievement.

Standards-Based Method. In this approach, teachers develop scoring guidelines for assignments to be included as components of the final grade. In the Hawaii Algebra Learning Project (**Sample A.31** in **Appendix A**—*Sampler*), all student journal entries are scored on a five point scale based on the quality of ideas, support for ideas, ability to communicate ideas to the reader and questions proposed for continued reflection. A "4" equals an "A" because the teachers decided that "4" reflects "A" work.

Figure 4.6 shows sample results for a particular student.

Figure 4.6
Sample Student Record for a Unit

Assessments Used		Score					Weighting
		0	1	2	3	4	
Problem-solving tasks	#1			X			44%
	#2			X			
	#3				X		
	#4					X	
	#5					X	
	#6					X	
	#7				X		
Communication tasks	#1			X			44%
	#2			X			
	#3					X	
	#4					X	
	#5					X	
	#6					X	
	#7					X	
Tests/application tasks	#1					X	12%
	#2					X	

If the student needed to know how he or she had performed for this period of time, the scores could be averaged. In this case, the average is 3.4 (4 scores worth 2 points each + 2 scores worth 3 points each + 10 scores worth 4 points each, all divided by 16 total scores). This could represent a B, based on cut-offs set by the teacher ahead of time. Or, some teachers may weigh later work more heavily than early work, determining it to be more reflective of what the student can do. The early scores of "2," for example, may reflect performance prior to understanding the criteria for good work. (See the next section for additional discussion of this option.)

Samples of work representing the four levels of quality could be available so parents could see how their child's work compared to higher or lower levels of work.

This standards-based grading approach is the one most consistent with the current national climate of developing performance standards and benchmarks. It may be time-consuming, at least at first, because of the need to define levels of achievement on tasks and help students understand the criteria for particular skills. However, as discussed in **Chapter 2**, such efforts are not empty exercises designed merely to grade students; they are also necessary components for improving instruction and assisting students in taking control of their own learning. For example, in the Hawaii Algebra Learning Project, students participate in developing the criteria for group presentations and journal prompt responses. These are "generalized criteria" to be applied across different kinds of problems and journal prompts. Because they are used repetitively, students internalize them and are able to accurately self assess across many assignments.

Final Thoughts About Grading

Regardless of the method used to convert scores to grades, there are several additional complexities to be considered.

Should grades include performances from early in the grading period or just those at the end? Students may perform more poorly on beginning assignments than they do on later assignments that assess the same skill. For example, as students continue to write answers to journal prompts, one would hope they would improve their mathematics communication skills (as in **Figure 4.6**). Thus, early scores for ability to communicate mathematics understanding may not present as accurate a picture of student ability as scores at the end of the grading period. So, it could well be argued that if parents are to see an accurate picture of their child's achievement level, only the mathematics communication scores for the journal entries at the end of the grading period should be considered. In other words, who cares if skill level is low at the beginning, as long as skills are good at the end?

Quality Issue Box

Notice that this approach requires a good description of "adequate skill" so that the teacher knows when the student is consistently performing at that level. And so we return to the necessity for good quality rubrics and training in their use. Also note that this approach might work better for "big" student outcomes such as problem solving, critical thinking, communication, and group collaboration than for content knowledge targets, once again illustrating the need to match targets to methods.

Should opportunities for revisions of work be allowed in certain skill areas (e.g., communication)? The development of certain skills (e.g., critiquing science articles, writing, oral presentations), is improved when students have the opportunity to revise products based on specific comments from the teacher. Grading these assignments the first time they are submitted may not contribute to the improvement of students' skills. In these situations, students might instead submit "finished" drafts to which teachers respond with suggestions for revision.

In a variation on this idea, students in the Hawaii Algebra Learning Project can receive a maximum score of 4 for their group presentation of the problem they solved. If the group or any individual within the group is unhappy with the score they received, they have the option to redo the discussion of the problem on paper and turn it in for an improved score. The final score is used in the grade.

By encouraging revision, the message to students is that improvement is valued. Also, practicing revision will tend to ensure that succeeding performances will be better, especially if the criteria for success are crystal clear.

Should student choice (e.g., goal-setting and assembling evidence of their own learning) be part of the grading process? Another choice facing teachers as students become more active partners in the learning process is the degree to which responsibility for proving the attainment of target outcomes is handed to students. Should a grade be a negotiated understanding between the teacher and student about the quality of the student's work based on evidence provided by the student?

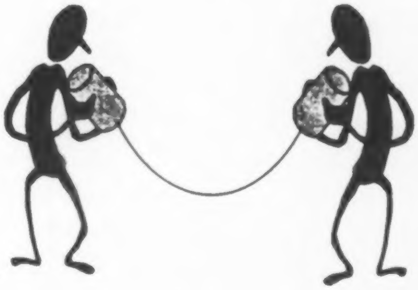
⊂Related Toolkit98 Chapters and Activities.

Related activities in this chapter are: Activity 4.6—*How To Convert Rubric Scores to Grades* and Activity 4.9—*Grading Scenarios*.

Students might be asked, for example, to develop a portfolio of work that they feel builds a case for their achievement and write a paper arguing for the grade they think they should earn. In this way, the grading process becomes part of the effort to develop students' capacity for self-assessment.

Reporting: Communicating About Student Learning

The expanding interest in alternative assessment has major implications for what is referred to as "reporting." Teachers who value the insights and student engagement that come from using alternative assessments such as open-ended performance tasks, exhibitions, and portfolios often worry about how to "report" student learning using such assessments. The word "reporting," itself, indicates a view of the communication between teacher and students, teacher



and parents, or school and community as essentially one-way. Some educators are calling for us to reframe our definition of assessment to address its essential role as a communication system—opening up conversations between students and their teachers, between students and parents, teachers and parents, and among students.

This view sees reporting as a communication system, a system for making meaning of children's learning and for taking action to improve teaching, curriculum, and learning. Thus, questions about "reporting" actually may be more about changing the way we communicate. (Please note that the grading suggestions in the previous section have the effect of changing the way we communicate progress to students and parents. They also have the effect of including students and parents in the assessment process.)

Options for Communicating About Student Learning

Frequently the issue of reporting is framed as redesigning report cards. As an alternative or supplement to current report cards, a number of schools are implementing reporting systems that more fully match their beliefs about the purposes of assessment in the overall picture of school learning. Some common characteristics of such alternatives are (1) an emphasis on communicating about complex skills, habits of mind, multiple intelligences and collaborative work, (2) using narratives or other alternatives to letter or number grades, and (3) reporting by themes or major concepts across subject areas. As with most innovations, the approaches taken to rethinking and redesigning communication about student learning vary widely—ranging from efforts to provide additions to existing categories reported, to wholly narrative and descriptive summaries, to student-led conferences that take the place of traditional report cards.

Related *Toolkit98* Chapters and Activities.

Further information and examples of report card formats are in Activities 4.4—*Won't Some Things Ever Change*? And 4.5—*Report Card Formats*. Additional examples of portfolios, senior projects, and developmental continuums can be found in Appendix A—*Sampler*.

Reference Box

A good source of other portfolio samples is: *Bibliography of Assessment Alternatives—Portfolios* (updated yearly), Northwest Regional Educational Laboratory, 503-275-9500.

The following ways to demonstrate student achievement are examples from among many possibilities. They represent several options most unlike traditional report cards.

Portfolios. Schools like Central Park East Secondary School in New York City use portfolios as demonstrations of learning. At Central Park West, students prepare 14 separate portfolios over an extended period of time to demonstrate their achievement. Students must "defend" their portfolios before a panel that includes a long-term student advisor, an adult selected by the student, another student, and someone with extensive knowledge of the portfolio's particular content. Rubrics for scoring portfolios spell out expectations and include such things as: viewpoint (the portfolio encompasses a wide knowledge base and is focused), connections (the whole portfolio is greater than the sum of the parts), evidence (there's credible and convincing evidence of achievement), voice (the portfolio is engaging), and conventions (the portfolio is intelligible).

Students determine their readiness to present their portfolios. They receive very focused feedback on what to do to improve the portfolio if it is not yet considered to have met expectations.

In Hawaii, one elementary school is using portfolios as one part of a reporting system to assist students to be self-confident risk takers and creative problem solvers, collaborative and capable in multiple dimensions. Parents, teachers, and students worked together to define these important learning targets. **Figure 4.7** is an example of part of a first grade report that students helped create.

Figure 4.7
Sample First Grade Report Card

D = Developing **P = Practicing** **I = Independent Application**	**Student**	**Teacher**	**Parent**
Creative Problem Solver Looks for different ways to solve a problem Knows where to find information			
Self-Confident Risk Taker Willing to participate in activities and discussions Willing to try new things Willing to try independently			
Collaborative Takes care of property Cooperates in a group Uses kind words Listens to others' ideas Finishes work on time			
Well-Rounded Willing to try a variety of activities			

Narratives. Some schools, such as Crow Island School in Winnetka, Illinois, use report cards that contain descriptions of achievement as well as portfolio evenings (in which parents are invited to discuss student portfolios) to engage students and parents in thinking and communicating about learning. An increasing number of schools that have begun using such narrative descriptions of learning are incorporating technology into their reporting systems. Juneau School District in Alaska is using computers to help teachers shorten the time it takes to write descriptions of student learning.

Developmental Continuums. Teachers in some schools have worked to describe the common characteristics of readers, writers, etc. at various developmental stages. Teachers might check or highlight the skills, behaviors, and knowledge that students currently exhibit and use this as an indicator of developmental level.

Multiple Sources. FIRST CLASS, an innovative middle school program in Rochester, New York, uses many forms of evidence of student learning in an ongoing reporting system, including report cards, interim interviews, written reflections and video/audiotaping, portfolios, conferences with families, presentations, exhibits, and standardized testing.

The report card for FIRST CLASS reports in four major areas: (1) engagement/class participation and involvement; (2) collaboration, cooperative learning groups, and working relationships with others; (3) independence and self-direction; and (4) performance/quality of work.

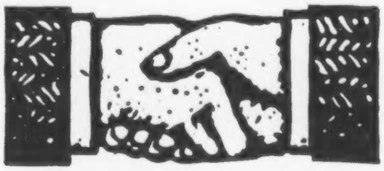
Reference Box:

Two nice, concise, useful booklets on student-involved conferences are: Nancy Little and John Allan, *Student-Led Teacher Parent Conferences*, 1988, Lugus Productions Limited, 416-322-5113; and Terri Austin, *Changing the View—Student-Led Parent Conferences*, 1994, Heinemann, (800) 541-2086.

Student-Involved Conferences. Schools in several districts in British Columbia have been using student-led and three-way conferences as alternatives to formal report cards. Three-way conferences take place when a child meets with his/her parents and teacher to demonstrate, show, and/or tell what has been learned. The meeting is "chaired" by the child. The child's learning strengths and areas of concern are discussed. Learning goals are set with the child, parent, and teacher. The teacher is available for questions, but the student has major responsibility for communicating about learning. As currently designed, brief written summaries of such conferences are permitted in place of more formal report cards several times a year. This option is part of a much larger system of ongoing communication among student, family, and teacher.

Conclusion

A central concern in any examination of reporting is the need for a good match among the learning that is desired (clear learning targets), the forms of assessment used to gather rich images of learning (target-method match), and the communication system for conversations about learning that lead to action and growth (grading and reporting). To gain insights into the appropriateness of various reporting options, ask yourself and others what *your* report card or report system communicates about the things that are valued within the school or system. What things are reported directly? Which things are not? How consistent is the reporting system with valued learning targets for students? Is information communicated in a way that is clear and promotes understanding? If so, then the reporting system works.



Activity 4.1

Weighty Questions

Purposes:

1. To develop sound grading practices that reflect valued student learning targets
2. To explore practical examples of how to incorporate alternative assessment information into report card grades
3. To examine current testing and grading practices in relation to desired outcomes for students

Uses:

This activity is at an advanced level. It should probably follow a series of activities that have familiarized participants with alternative assessment. It assumes that they will be familiar with a variety of alternative assessments (**Chapter 3** text), the idea of matching assessments to valued student learning targets (**Chapter 1** text and **Activities 1.2**—*Clear Targets* and **1.7**—*Target Method Match*), and how assessment fits into standards-based education (**Chapter 1** text or **Activities 1.1**—*Changing Assessment Practices*, **1.6**—*A Comparison of Multiple-Choice and Performance Assessment*, or **1.12**—*Assessment Principles*). The reading material in **Chapter 4** could be distributed and read prior to this session. **Activities 4.2**—*Putting Grading and Reporting Questions in Perspective* or **4.3**—*Grading Jigsaw* are also good stage setters.

Rationale:

Teachers may see the need for alternative assessments such as journals, portfolios, projects, etc., but have a difficult time envisioning how to make them fit with district demands that the assignment of grades be based on a percent system (e.g., 93 to 100%=A). This activity supports teachers in articulating the choices they have made in their current testing and grading processes and making decisions about what they would like to change.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead A4.1,O1**—*Weighty Questions Purposes*
- **Handout A4.1,H1**—*What Student Learning Targets Do/Should Grades Assess?*; **Sample A.31**—*Hawaii Algebra Grading Process* from **Appendix A**—*Sampler*; **Chapter 4** text

Time Required:

90 minutes to 2 hours; extension requires an additional 60 minutes

Facilitator's Notes:

A. Basic Activity

1. (2 minutes) Introduce the purposes of the activity, as needed, using **Overhead A4.1,O1**—*Weighty Questions Purposes.*

2. (30 minutes) Review the kinds of decisions (outlined in **Chapter 4**) to be made in giving report card grades, perhaps using **Sample A.31**—*Hawaii Algebra Grading Process* (from **Appendix A**—*Sampler*) as an illustration.

3. (10-20 minutes) Introduce the activity by telling participants that they will discuss the decisions about grading they have made. Pass out **Handout A4.1,H1**—*What Student Learning Targets Do/Should Grades Assess*? Ask participants to *individually* complete the first three columns for a course or subject area they teach, thinking about the current or last grading period. For example, a science teacher might list quizzes and tests under "assessment methods," indicate that the "knowledge/skills assessed" by the tests used were recall, conceptual understanding, and application, and record that this assessment method accounted for 60 percent of a student's report card grade. They might list projects as the next assessment method used, and under "knowledge/skills assessed" enter plan, research, synthesize, create a product, present information, develop conceptual understanding, and defend a point of view. They might say projects counted as 20 percent of the final grade. And so on.

4. (15-30 minutes) After participants have had time to complete the information individually, ask them to share their responses in small groups. Each group should to try to come to some kind of consensus about the desired outcomes and weighting for the course/grade they teach. They should complete the last two columns to report out to the whole group. The reporting might prompt a

discussion of whether or not teachers in a school should try to come to agreement about important student learning targets to be assessed in the grading process.

5. (15-30 minutes) Finally, ask each group to refer to the pages in **Chapter 4** describing how grades might be assigned (percent averaging, total point, standard-based, other options) and list the methods used by group members. The methods used could be other than those described in the chapter. Groups should report back to the whole group with discussions of any questions, helpful suggestions, etc.

B. Extensions (60 minutes)

1. As a follow-up activity, teachers could be asked to bring a test they have recently used and work in groups to list the kinds of student learning targets assessed by each item (e.g., recall, analysis, comparison, inference, evaluation). The point of this analysis would be to (a) find out if the tests actually do assess the student learning targets valued by faculty, and (b) to continue conversations among teachers in a school about what student learning targets are valued and how assessments might be brought more in line with these valued learning targets.

2. Participants might also be asked to discuss:

   a. The extent to which they modify grades for student effort, generating a list of the pluses and minuses of doing so.

      It is desirable that the conclusion is that minuses outweigh the pluses because when the waters are muddied with "effort" (which tends to be assessed and defined in different ways by different teachers), we don't have a strong communication system about student achievement. Nor does promising "A's" and threatening "F's" motivate students. These notions are supported by current research and expert opinion (see the articles in **Appendix C**—*Articles*). At this point you could also ask teachers, "How else might we strengthen student motivation?"

   b. The extent to which they modify grades in consideration of student ability and the pluses and minuses of doing so.

      It is desirable that any modification of grades for "ability" be done in advance. Factoring in ability after the fact—after the work is done—has the great potential to backfire by doing harm to student self-concept and motivation. Other thoughts:

      - Base modifications on the learning targets to be achieved, not on lowering the "%'s" for students

      - Base modifications on past achievement, not ability, which we can only guess at.

Weighty Questions

Purposes:

1. To develop sound grading practices that reflect valued student learning targets
2. To explore practical examples of how to incorporate alternative assessments into report card grades
3. To examine current testing and grading practices in relation to desired learning targets for students

Overhead A4.1,O1

What Student Learning Targets Do/Should Grades Assess?

List assessment methods reflected in gradebook (e.g., paper and pencil tests/quizzes, homework, journals portfolios, projects, performance tasks)	What knowledge/skills (outcomes) are assessed by each method	Estimate current weighting (% of total grade)	Reconsider outcomes: List knowledge/skills that should be assessed	Enter desired weighting for each outcome area (% of total grade)
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				

Handout A4.1,H1

Activity 4.2
Putting Grading and Reporting Questions in Perspective

Goals:

1. To think about issues surrounding grading and reporting
2. To set the stage for ongoing discussions about grading

Uses:

This is an intermediate level activity. It is a good one to use to kick off discussions about grading and reporting. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); and (b) understanding of how assessment can be integrated with instruction (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*).

Rationale:

The overall instructional philosophy in *Toolkit98* is that everyone—students, teachers, and adults of all kinds—learn best when they are involved in a hands-on experience and then draw their own conclusions. This does not, however, mean that we leave teachers totally up to their own devices. Teachers need material to react to and try out. This is the philosophy behind this activity.

In our experience, grading and reporting strikes a deep emotional chord in teachers. Unstructured discussions can frequently go nowhere. Therefore, this activity is based on reading and reacting to a somewhat controversial article about grading by Alfie Kohn (see Article #1 in **Appendix C**—*Articles*). The goal is *not* to make every teacher embrace Kohn's ideas. The goal is to remind teachers of the times they feel uncomfortable about practices in grading and reporting, and give them the opportunity to discuss what they would like to do about it.

Materials Needed:

- Overhead projector, screen, blank transparencies, and transparency pens
- Flip chart sheets and pens—one per group of 4-6
- **Overheads: A4.2,O1**—*Putting Grading...Purposes and Agenda* and **A4.2,O2**—*What are the 3 biggest questions*...(Handouts **A4.2,H2** and **A4.2,H3** can also be made into overheads.)
- **Handouts: Handout A4.2H1**—*Instructions for 3 Biggest Questions,* **A4.2,H2**—*Grading*.... **A4.2,H3**—*Points made by previous groups*; one per person

Time:

75-90 minutes

Facilitator's Notes:

1. (2 minutes) Go over the goals and agenda for the segment. (See **Overhead A4.2,O1**—*Putting Grading and Reporting Questions in Perspective Purposes and Agenda.*) If this is the opening activity to a further sequence of activities, tell participants that this activity will add a common experience and discussion to the mix before addressing specific grading issues.

2. (20-25 minutes) Participants will first individually brainstorm their three biggest questions about grading and reporting. Then they'll choose a recorder for their group and come to agreement about the group's biggest three questions. The recorder will write the questions on chart pack paper and post them. Finally, there will be a "museum tour" in which groups circulate to the questions posted by other groups to see how many questions echo their own. Instructions and space to write down questions is provided on **Handout A4.2,H1**—*Instructions for 3 Biggest Questions.*

   Overhead A4.2,O2—*What are the 3 biggest questions*...can be used to stimulate thinking if individuals and groups have trouble getting started.

3. (45-60 minutes) In this segment, participants are going to read Alfie Kohn's article (in **Appendix C**), identify where in Kohn's three levels of concern about grading and reporting their questions fall, and think about what ideas they like and dislike in Kohn's paper.

   a. (5 minutes) Providing an introduction to Kohn's paper before reading begins enables participants to focus faster on the important elements and keep the paper in perspective. Points to include in an overview:

(1) Kohn's paper is excellent in providing the big picture about grading. It's a must read. This doesn't mean, however, that we're necessarily promoting all his ideas. The purpose of the paper is to raise issues not dictate solutions. It provides a common base from which to discuss issues.

At the end of reading, participants will be asked to identify what ideas they liked, what ideas they felt uncomfortable about, and what ideas they found "interesting" (that is, they would need to think about them more before forming an opinion).

(2) Kohn will discuss three levels of concern about grading. Level I concerns deal primarily with the "how to" questions such as "How do I combine numbers to form grades?" Level II concerns go beyond traditional sources of information to ask whether traditional grading is really necessary or useful for assessing students' performance. Are there better ways to represent student learning than grades? At this level, also, teachers might not question the need to grade, but feel that grades need to be based on sounder evidence; so, they begin to want to incorporate "alternative assessment" information into their grades. Level III concerns about grading focus on questions of, "Why do we grade?" and "Why do we assess?" *"No matter how elaborate or carefully designed an assessment strategy may be, the result will not be constructive if our reason for wanting to know is itself objectionable."* (Kohn, p. 2)

At the end of reading participants will identify the levels associated with the questions they asked about grading and reporting.

(Knowing the level of questions of a group also assists in knowing where to begin, or the approach to take, in training or discussion about grades.

b. (30-40 minutes) Read Kohn's paper and individually (1) decide on the level of their own questions and (2) identify "likes," "dislikes," and "interesting's." **Handout A4.2,H2**—*Grading*.... provides space for notes. Remind groups that they should leave 10 minutes at the end to discuss: levels of their questions and their "likes," "dislikes," and "interesting's."

c. (10-15 minutes) The debrief has two parts. First ask for a show of hands about levels. Then have a quick review of "likes," "dislikes" and "interesting's." A good technique for moving things along is to "stand and share." For each category (likes, dislikes and interesting) ask one person to stand from each table if they have something to share. If any of the people say what they were going to say they sit down.

(Comments from past workshops are provided in **Handout H3**—*Points Made by Previous Groups*. Other comments: Likes—intrinsic motivation; having grades represent achievement only. Dislikes—doesn't include the rationale for grading; doesn't mention the need for multiple sources of information. Interesting—Kohn says evidence "seems to suggest" rather than offering firm research evidence.)

This activity should not lead participants to conclude that they shouldn't have to grade. Grading is a political reality that will be with us for some time to come. The conclusion should be that, since we have to grade, let's make the process as meaningful as possible. For example, encouraging grades to be a shorthand for well-defined, standards-based achievement levels; thinking of ways grading practices can encourage learning and support student success; and avoiding undesirable side effects.

4. (5 minutes) You might end the session by asking participants to write down one idea that they are going to try, or one piece of information they want to remember.

Putting Grading and Reporting Questions in Perspective

Purposes:

1. To think about grading issues
2. To set the stage for more to come

Agenda:

- Review and discuss three biggest grading questions
- Read a paper by Alfie Kohn
- Relate your questions to Kohn's

Overhead A4.2,O1

What are the 3 biggest questions about grading and reporting that you hear from your colleagues?

For example:

- How do I incorporate "authentic" assessment into my grades?
- How do I grade special education kids? Do I hold all students to the same standard?
- Our report card doesn't seem to reflect what we're trying to accomplish with students.
- Why do we need to grade? We see students every day, so we know how well they're doing.
- We're required by law to demonstrate how well classroom activities help students move toward the state's student learning targets, but my grades don't seem to match to the state standards.
- How are grading criteria developed?
- How do I write grading criteria (standards) so parents and students understand?

Overhead A4.2,O2

Instructions for "3 Biggest Questions"

1. Take a few moments to write your own personal questions in the space below.
2. Choose a recorder for your group. Come to agreement on the three biggest questions, write them on chart pack paper, and post them on the wall.
3. When everyone has their questions posted, take a few minutes to see how others' questions relate to yours.

Space to Write Down Your Questions

Handout A4.2,H1

"Grading: The Issue Is Not How But Why" Discussion Questions

1. Where in Kohn's three levels of concern about grading and reporting do your questions fall?

   Level I: How to combine #s

   Level II: Beyond traditional

   Level III: Why grade? Why assess?

2. List at least one thing from the article in each category:

- I like...
- I don't like/this makes me nervous...
- This is interesting; I want to think more about it

Handout A4.2H2, p.2

570

Points made by previous

groups:

- Grades are only as good as the assessments on which they are based
- The only legitimate purpose for grades is feedback on achievement and a single letter grade doesn't do this very well
- There exist better ways to provide feedback than grades
- We should be addressing more fundamental questions than "how do we manipulate numbers?"
- Most of the questions we hear about grading and reporting seem to be at Level II
- Many times questions that manifest themselves as Level I or II concerns really require Level III answers
- If we must grade, we need to make the process as communicative as possible and avoid possible negative side effects

Activity 4.3
Grading Jigsaw

Purpose:

To think about grading and reporting issues

Uses:

This is an intermediate level activity. It is a good one to use to kick off discussions about grading and reporting. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** Text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); and (b) understanding of how assessment can be integrated with instruction (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*).

Rationale for the Activity:

Reading large numbers of articles is time-consuming. A jigsaw allows people to work together to read and discuss many articles in a short period of time. Each person reads a single article. Then people who read different articles get together to discuss the content of what they read.

There is no single "right" approach to grading and reporting. We have chosen articles (see **Appendix C**—*Articles*) that describe what others have done and to raise issues for discussion. We have based our selection of articles on Alfie Kohn's article (#1 in **Appendix C**) about teachers' level of concern regarding grading. The first, and most basic, level of concern has to do mainly with logistics—"how to combine all the information collected to form a final grade." The second level of concern begins to question why we need to grade. Might there be better ways to show student achievement? Teachers at the second level might also start questioning traditional ways of collecting information. Teachers don't question the need to grade, but want to expand the types of information collected to form a grade. (This is where many discussions of "alternative assessment" begin.) At the third level of concern teachers begin to challenge the notion of grading and assessing altogether. At this level the question is, "Why do we assess or grade at all? Are we accomplishing what we think we are?" The papers chosen for this jigsaw address mostly the second and third levels of concern.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens, flip chart, and pens
- Set of articles on grading (**Appendix C**—*Articles*). You can either copy enough of the articles so that each participant only receives the single paper he or she will use in the jigsaw, or you can copy an entire set for each participant. (Copy the entire set for each person if participants do not come from the same school or organization.) Suggestions for which papers to use are in the **Appendix C**—*Articles* section. (All papers have been reprinted with permission of the author or publisher. Permission has been given to copy them for use in professional development settings.)
- **Overhead A4.3,O1**—*Grading Jigsaw Purposes*, **A4.3,O2**—*Grading Jigsaw Discussion Questions*
- **Handout A4.3,H1**—*Grading Jigsaw Discussion Questions* (can also be used as an overhead)

Time Required:

90 minutes-2¼ hours

Facilitator's Notes:

1. Set-Up

   a. First, an overview of the jigsaw process. A jigsaw is designed to share large amounts of information in a short period of time by having participants become experts on a piece of information and then share it with others. The jigsaw begins by dividing participants into small groups, each of which will read and discuss a different article, thereby allowing group members to become "experts" on that article. Then the groups reform so that there is, in each new group, an expert on each paper. Each person shares their expertise with the new group. So, everyone learns about all the papers without having to personally read them.

   For example, let's say you have 24 people and six articles; four people will read each of the six articles. Each group of 4 discusses their own article to become the "expert" on that article. Then the participants regroup so that each new group has an "expert" on each of the articles. In our example, we started out with 24 people, four reading each of the six articles. When people regroup, there will be four groups of six people; each group has an "expert" on each paper.

   b. You need to decide which articles to copy. We chose some of the articles for groups that are just starting their discussion of grading and report changes; we chose other articles for more experienced groups. These suggestions are in the Index to **Appendix C**—*Articles*. You may take our suggestions or decide for yourself which papers would be of most interest for *your* group.

c. Decide how many copies to make of each article. If your budget is tight, all you really need is one paper per person—make an equal number of copies of each article you have chosen to use, but any single person only gets one. In our example, since we have 24 people, we would make four copies of each of six articles. (Of course, you could always make a copy of *all* the chosen articles for each participant. Even though participants will only need a single paper during the jigsaw, they might want to have a complete set.)

d. Put the copies of each article on a separate table. For our example, we would have six tables (one for each of the six articles); each table would have four copies of the same article. Letter the papers at each table, beginning at A and continuing until all the papers have a letter. (In our example, we have four papers at each table; we'd letter them A-D.) You do this to make it easy for groups to rearrange themselves later—after the "same article" groups discuss their article, you have all the A's, B's, C's, etc. get together; this gives you one "expert" in each group.

2. (5 minutes) Discuss the purpose of the activity using **Overhead A4.3,O1**—*Grading Jigsaw Purposes*.

3. (30-45 minutes) Spread participants around so that the same number of people are at each table. Ask participants to read the paper on their table and *come to agreement* about the most important points—both what they liked and what they didn't like. You can use the questions on **Handout A4.3,H1**—*Grading Jigsaw Discussion Questions* to focus the discussion. The questions are:

   - Why do we grade?
   - What are the issues surrounding grading and report cards?
   - Why are some people changing grading and report cards?
   - What are the alternatives to grading to more accurately reflect student achievement and progress in ways students and parents can understand?
   - What report card formats have others tried?

4. (30-45 minutes) Now ask participants to move into "mixed paper" groups. You can do this by saying, "All the A's go to the table in the back, all the B's meet at the overhead projector, ..." etc.

   Each expert spends 5-10 minutes explaining the most important points of his or her paper to the others in the group. This will take 30-45 minutes depending on the number of articles.

5. (30 minutes) Reconvene the big group for a final discussion. Summarize the points from the individual discussion groups. Major conclusions should include:

   - The primary purpose for grading should be communication about student achievement.

- Report cards should use the language of the message receiver (students and parents).
- There are several formats, besides grades, that communicate student achievement.

6. (5-15 minutes) The large-group discussion can be followed by a discussion of either what participants might like to try in their own classrooms or how the current grading policy in their district compares to the concensus points from the final large group discussion.

Grading Jigsaw

Purpose:

To think about grading and reporting issues

Overhead A4.3,O1

576

Grading Jigsaw
Discussion Questions

- Why do we grade?
- What are the issues surrounding grading and report cards?
- Why are some people changing grading and report cards?
- What are the alternatives to grading to more accurately reflect student achievement in ways students and parents can understand?
- What report card formats have others tried?

Overhead A4.3,O2

Grading Jigsaw—
Discussion Questions

Article Title:

Question	What we liked from this article	What we didn't like from this article
Why do we grade?		
What are the issues surrounding grading and report cards?		
Why are some people changing grading and report cards right now		
What are the alternatives to grading to adequately communicate student achievement and progress?		
What report card formats have others tried?		

Handout A4.3.H1

Activity 4.4

Won't Some Things Ever Change?

Purposes:

1. To compare and contrast traditional reporting formats for student report cards over several generations
2. To critically examine the information conveyed about what a student knows and is able to do by traditional reporting formats

Uses:

This is an intermediate level activity. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); (b) understanding of how assessment can be integrated with instruction (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*); and (c) **Activity 4.2**—*Putting Grading and Reporting Issues in Perspective* to set the context for examining report card designs.

Rationale:

The traditional letter grade report card format has been used in the United States educational system for generations. Although some changes are noticeable, such report cards do not seem to reflect recent major instructional and assessment shifts that have resulted from new knowledge about the way people learn.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens

- **Overheads A4.4,O1**—*Won't Some Things Ever Change Purposes;* **A4.4,O2**—*Compare and Contrast*

- **Handouts A4.4,H1**—*Grade 6 Report Card, 1916: Essie;* **A4.4,H2**—*Grade 12 Report Card, 1943-44: Oscar*; **A4.4,H3**—*Grade 10 Report Card, 1965-66: Jackie;* **A4.4,H4**—*Grade 3 Report Card, 1994-95: Myrth* (make each handout a different color)

Time Required:

30-40 minutes

Facilitator Notes:

1. (2 minutes) Use **Overhead A4.4,O1**—*Won't Some Things Ever Change Purposes*, as needed, to introduce the purposes of the activity.

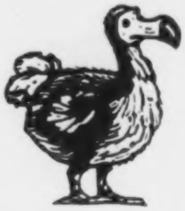
2. (15-10 minutes) Use **Overhead A4.4,O2**—*Compare and Contrast* to guide small groups of 3-5 participants in comparing and contrasting the report card formats exhibited in **Handouts A4.4,H1, A4.4,H2, A4.4,H3,** and **A4.4,H4**.

   - How are they alike?
   - How are they different?
   - What does each report card convey about the student's knowledge, skills, and understanding?
   - Who is the target audience? Would that audience understand the message sent?
   - What does each report card convey about teaching and learning?

3. (10 minutes) Have each group report out highlights of their discussions. Emphasize that, today, the main purposes for grading and reporting are:

   - Encourage learning and support student success
   - Avoid undesirable side effects
   - Be meaningful/understandable to the recipient of the information (students and teachers)
   - Have the primary purpose of accurately communicating about student achievement

Won't Some Things Ever Change?



Purposes:

1. To compare and contrast traditional reporting formats for report cards over several generations

2. To critically examine the information conveyed in traditional reporting formats

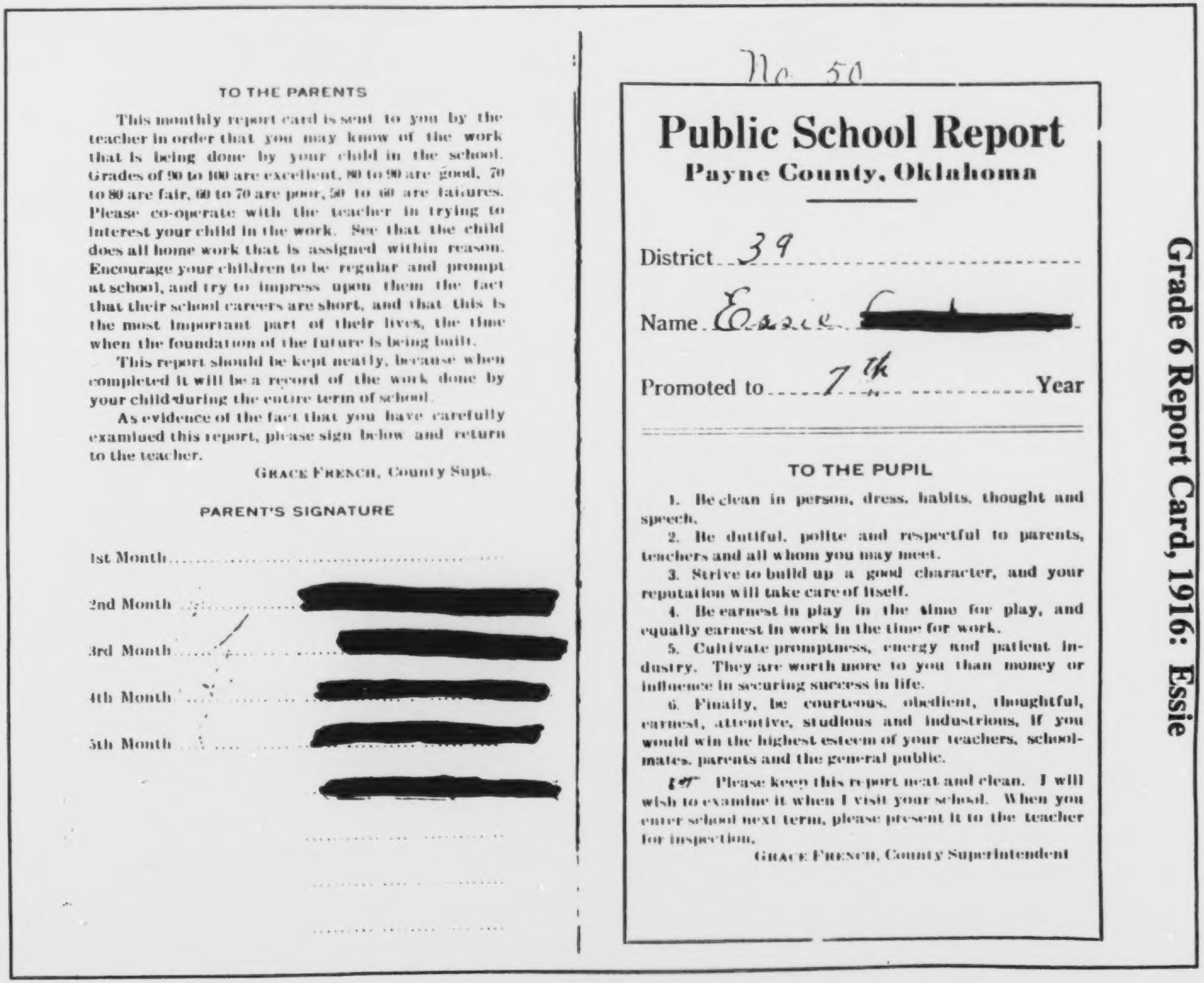
Overhead A4.4,O1

Compare and Contrast:

- How are they alike?
- How are they different?
- What does each report card convey about the student's knowledge, skills, and understanding?
- Who is the target audience? Would that audience understand the message sent?
- What does each report card convey about teaching and learning?

Overhead A4.4,O2

Grade 6 Report Card, 1916: Essie



TO THE PARENTS

This monthly report card is sent to you by the teacher in order that you may know of the work that is being done by your child in the school. Grades of 90 to 100 are excellent, 80 to 90 are good, 70 to 80 are fair, 60 to 70 are poor, 50 to 60 are failures. Please co-operate with the teacher in trying to interest your child in the work. See that the child does all home work that is assigned within reason. Encourage your children to be regular and prompt at school, and try to impress upon them the fact that their school careers are short, and that this is the most important part of their lives, the time when the foundation of the future is being built.

This report should be kept neatly, because when completed it will be a record of the work done by your child during the entire term of school.

As evidence of the fact that you have carefully examined this report, please sign below and return to the teacher.

GRACE FRENCH, County Supt.

PARENT'S SIGNATURE

1st Month ....

2nd Month [redacted]

3rd Month [redacted]

4th Month [redacted]

5th Month [redacted]

[redacted]

No 50

Public School Report

Payne County, Oklahoma

District 39

Name Essie [redacted]

Promoted to 7th Year

TO THE PUPIL

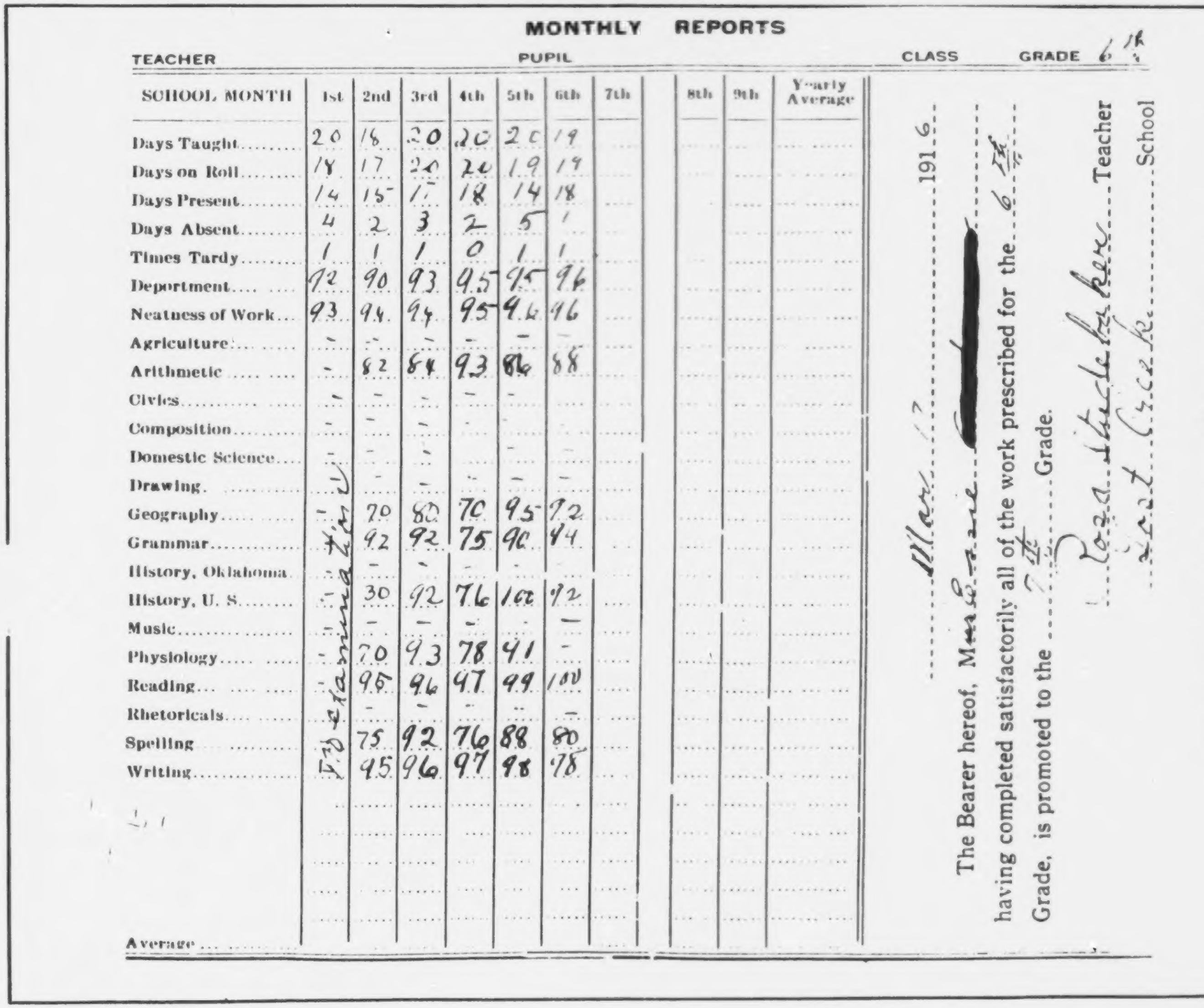
1. Be clean in person, dress, habits, thought and speech.
2. Be dutiful, polite and respectful to parents, teachers and all whom you may meet.
3. Strive to build up a good character, and your reputation will take care of itself.
4. Be earnest in play in the time for play, and equally earnest in work in the time for work.
5. Cultivate promptness, energy and patient industry. They are worth more to you than money or influence in securing success in life.
6. Finally, be courteous, obedient, thoughtful, earnest, attentive, studious and industrious, if you would win the highest esteem of your teachers, schoolmates, parents and the general public.

☞ Please keep this report neat and clean. I will wish to examine it when I visit your school. When you enter school next term, please present it to the teacher for inspection.

GRACE FRENCH, County Superintendent

BEST COPY AVAILABLE

Grade 6 Report Card, 1916: Essie



MONTHLY REPORTS

TEACHER PUPIL CLASS GRADE 6th

SCHOOL MONTH	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Yearly Average
Days Taught	20	18	20	20	20	19				
Days on Roll	18	17	20	20	19	19				
Days Present	14	15	17	18	14	18				
Days Absent	4	2	3	2	5	1				
Times Tardy	1	1	1	0	1	1				
Deportment	92	90	93	95	95	96				
Neatness of Work	93	94	94	95	96	96				
Agriculture	-	-	-	-	-	-				
Arithmetic	-	82	84	93	86	88				
Civics	-	-	-	-	-					
Composition	-	-	-	-	-					
Domestic Science	-	-	-		-	-				
Drawing	Examination	-	-	-	-	-				
Geography	-	70	80	70	95	92				
Grammar		92	92	75	90	94				
History, Oklahoma	-	-	-	-	-					
History, U. S.	-	30	92	76	100	92				
Music	-	-	-	-	-	-				
Physiology	-	70	93	78	91	-				
Reading	-	95	96	97	99	100				
Rhetoricals		-	-	-	-	-				
Spelling		75	92	76	88	80				
Writing		95	96	97	98	78				
Average										

Mar 17 1916

The Bearer hereof, M Essie [redacted] having completed satisfactorily all of the work prescribed for the 6th Grade, is promoted to the 7th Grade.

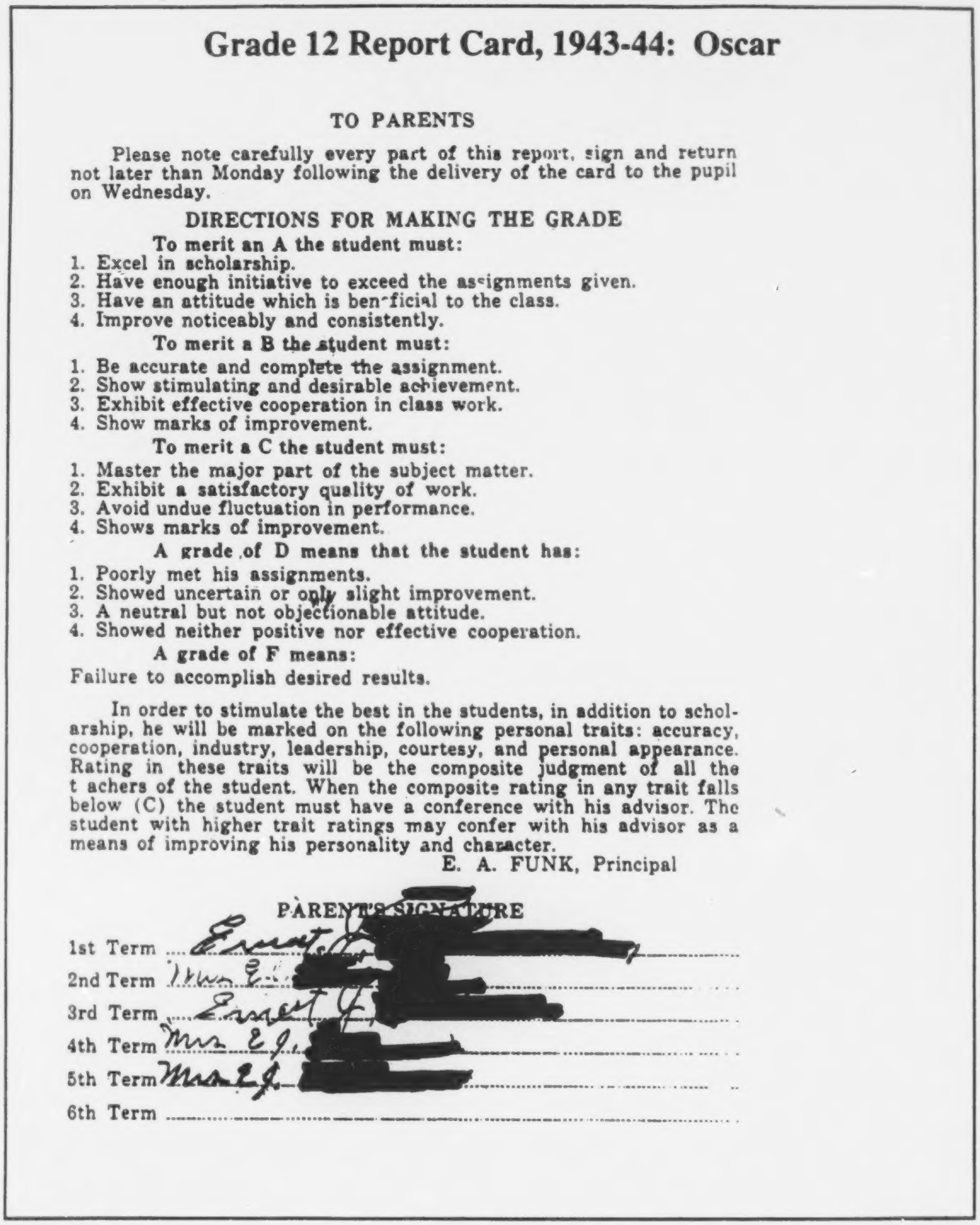
Roza Stucklebaker Teacher

East Creek School

BEST COPY AVAILABLE

585

Handout A4.4, H1, p. 2



Grade 12 Report Card, 1943-44: Oscar

TO PARENTS

Please note carefully every part of this report, sign and return not later than Monday following the delivery of the card to the pupil on Wednesday.

DIRECTIONS FOR MAKING THE GRADE

To merit an A the student must:

1. Excel in scholarship.
2. Have enough initiative to exceed the assignments given.
3. Have an attitude which is beneficial to the class.
4. Improve noticeably and consistently.

To merit a B the student must:

1. Be accurate and complete the assignment.
2. Show stimulating and desirable achievement.
3. Exhibit effective cooperation in class work.
4. Show marks of improvement.

To merit a C the student must:

1. Master the major part of the subject matter.
2. Exhibit a satisfactory quality of work.
3. Avoid undue fluctuation in performance.
4. Shows marks of improvement.

A grade of D means that the student has:

1. Poorly met his assignments.
2. Showed uncertain or only slight improvement.
3. A neutral but not objectionable attitude.
4. Showed neither positive nor effective cooperation.

A grade of F means:

Failure to accomplish desired results.

In order to stimulate the best in the students, in addition to scholarship, he will be marked on the following personal traits: accuracy, cooperation, industry, leadership, courtesy, and personal appearance. Rating in these traits will be the composite judgment of all the teachers of the student. When the composite rating in any trait falls below (C) the student must have a conference with his advisor. The student with higher trait ratings may confer with his advisor as a means of improving his personality and character.

E. A. FUNK, Principal

PARENT'S SIGNATURE

1st Term Ernest [redacted]
2nd Term Mrs. E. [redacted]
3rd Term Ernest J. [redacted]
4th Term Mrs. E. J. [redacted]
5th Term Mrs. E. J. [redacted]
6th Term ....

Handout A4.4, H2, p. 1

536

BEST COPY AVAILABLE

Grade 12 Report Card, 1943-44: Oscar

Anne Hawley
Distributing Teacher

SENIOR HIGH SCHOOL
ARKANSAS CITY, KANSAS
13

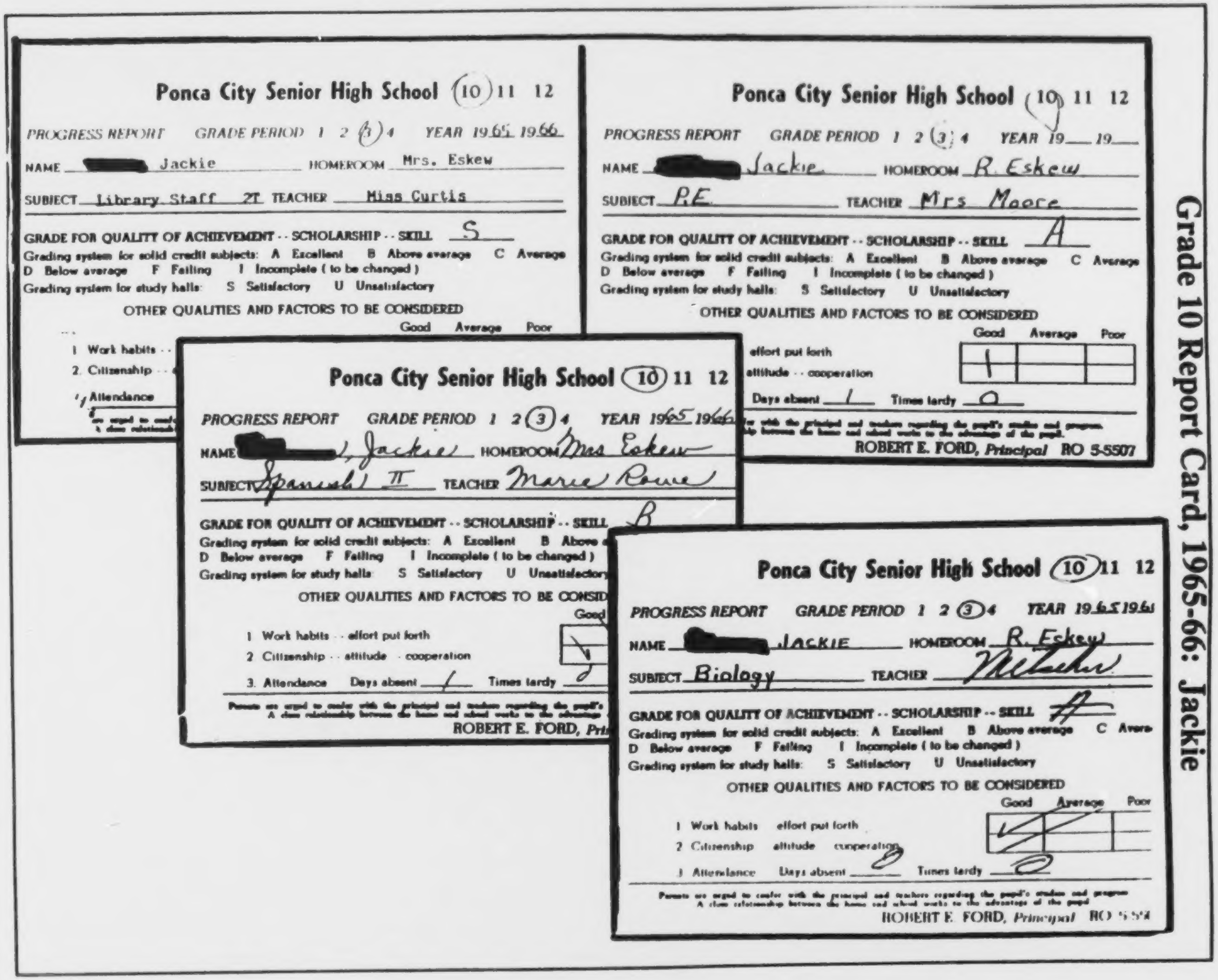
Page 35

REPORT OF [redacted] Oscar J. FOR 19 43 -19 44

SUBJECTS	1st Term			2nd Term			3rd Term			1st Sem. Average	4th Term			5th Term			6th Term			2nd Sem. Average	Units Credit
	Grade	Periods Absent	Times Tardy	Grade.	Periods Absent	Times Tardy	Grade	Periods Absent	Times Tardy		Grade	Periods Absent	Times Tardy	Grade	Periods Absent	Times Tardy	Grade	Periods Absent	Times Tardy		
Pub. Speak	A	1	1	A	2		A			A	B			A	3		B	4		B	1
Const	B	1		B	2		B			B											1/2
Prob											B			B	2		B	3		B	1/2
Physics	B	2		B	3		B			B	B			D	2		B	2		B	1
Adv. Alg	C	1		C	4	2	B			C											1/2
Sol. Geom											B			C	2	1	B	3		B	1/2
Drawing											B			B	2		B			B	1/2
Study		1										1			2			3			
Total Abs		6	1		11	2						1			13	1		15			

BEST COPY AVAILABLE

Grade 10 Report Card, 1965-66: Jackie



Ponca City Senior High School [✓ 10] 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 1965 1966

NAME [redacted] Jackie HOMEROOM Mrs. Eskew

SUBJECT Library Staff 2T TEACHER Miss Curtis

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL S

Grading system for solid credit subjects: A Excellent B Above average C Average D Below average F Failing I Incomplete (to be changed)

Grading system for study halls: S Satisfactory U Unsatisfactory

OTHER QUALITIES AND FACTORS TO BE CONSIDERED

Good Average Poor

1 Work habits - -

2. Citizenship - -

3. Attendance

Ponca City Senior High School [✓ 10] 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 19__ 19__

NAME [redacted] Jackie HOMEROOM R. Eskew

SUBJECT P.E. TEACHER Mrs Moore

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL A

Grading system for solid credit subjects: A Excellent B Above average C Average D Below average F Failing I Incomplete (to be changed)

Grading system for study halls: S Satisfactory U Unsatisfactory

OTHER QUALITIES AND FACTORS TO BE CONSIDERED

	Good	Average	Poor
effort put forth	✓		
attitude - - cooperation	✓		

Days absent 1 Times tardy 0

ROBERT E. FORD, *Principal* RO 5-5507

Ponca City Senior High School [✓ 10] 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 1965 1966

NAME [redacted], Jackie HOMEROOM Mrs Eskew

SUBJECT Spanish II TEACHER Marie Rowe

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL B

Grading system for solid credit subjects: A Excellent B Above

D Below average F Failing I Incomplete (to be changed)

Grading system for study halls: S Satisfactory U Unsatisfactory

OTHER QUALITIES AND FACTORS TO BE CONSID

Good

1 Work habits - - effort put forth ✓

2 Citizenship - - attitude - cooperation ✓

3. Attendance Days absent 1 Times tardy 0

ROBERT E. FORD, *Pri*

Ponca City Senior High School [✓ 10] 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 1965 1966

NAME [redacted] JACKIE HOMEROOM R. Eskew

SUBJECT Biology TEACHER

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL A

Grading system for solid credit subjects: A Excellent B Above average C Avera

D Below average F Failing I Incomplete (to be changed)

Grading system for study halls: S Satisfactory U Unsatisfactory

OTHER QUALITIES AND FACTORS TO BE CONSIDERED

	Good	Average	Poor
1 Work habits effort put forth	✓		
2 Citizenship attitude cooperation	✓		

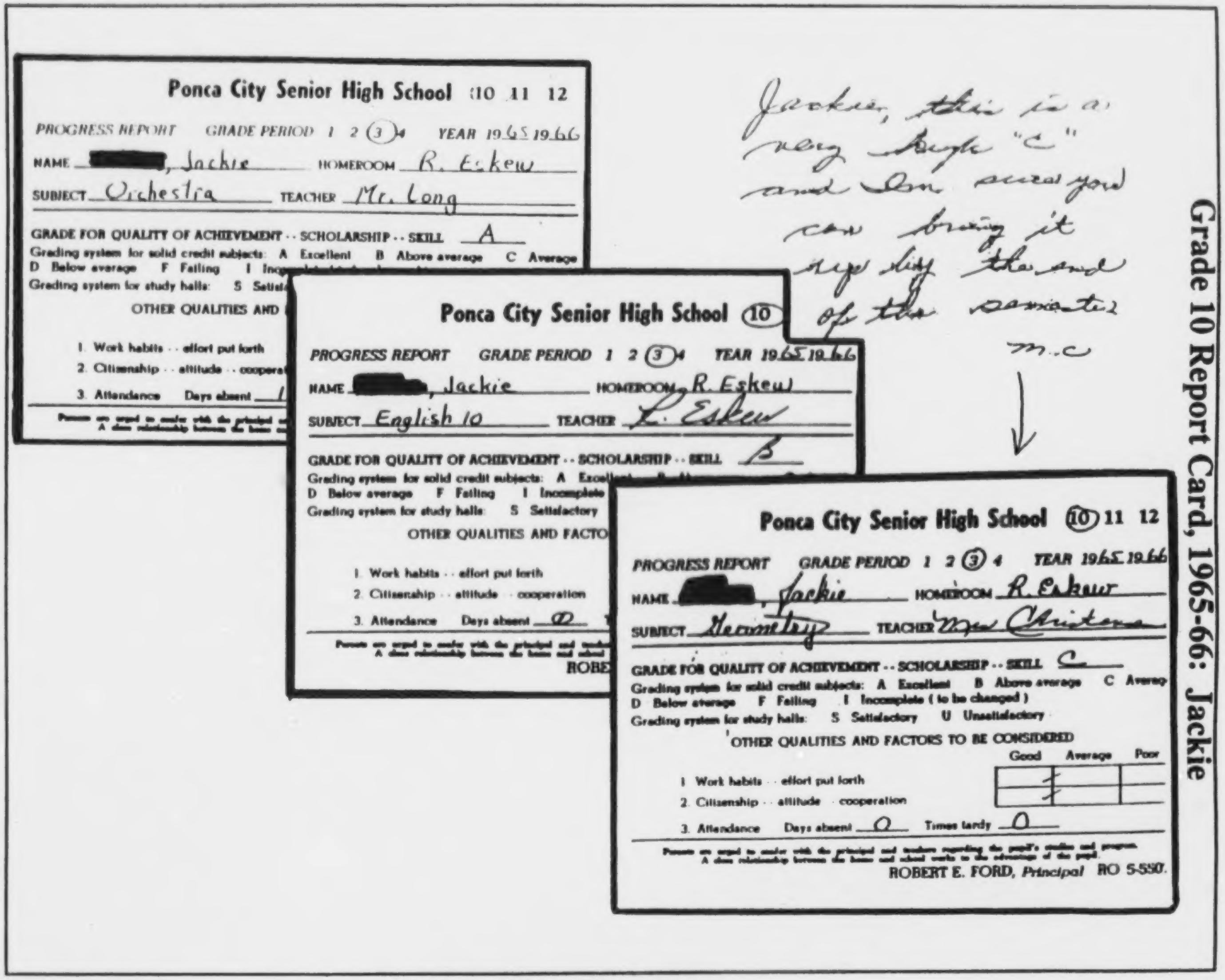
3 Attendance Days absent 0 Times tardy 0

Parents are urged to confer with the principal and teachers regarding the pupil's studies and progress. A close relationship between the home and school works to the advantage of the pupil.

ROBERT E. FORD, *Principal* RO 5-55

BEST COPY AVAILABLE

Grade 10 Report Card, 1965-66: Jackie



Ponca City Senior High School 10 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 19 65 19 66

NAME [redacted], Jackie HOMEROOM R. Eskew

SUBJECT Orchestra TEACHER Mr. Long

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL A

Grading system for solid credit subjects: A Excellent B Above average C Average

D Below average F Failing I Inc

Grading system for study halls: S Satisf

OTHER QUALITIES AND

1. Work habits - - effort put forth
2. Citizenship - - attitude - - cooperat
3. Attendance Days absent 1

Ponca City Senior High School [✓ 10]

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 19 65 19 66

NAME [redacted], Jackie HOMEROOM R. Eskew

SUBJECT English 10 TEACHER L. Eskew

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL B

Grading system for solid credit subjects: A Excell

D Below average F Failing I Incomplete

Grading system for study halls: S Satisfactory

OTHER QUALITIES AND FACTO

1. Work habits - - effort put forth
2. Citizenship - - attitude - - cooperation
3. Attendance Days absent 0

ROBE

Jackie, this is a very high "C" and I'm sure you can bring it up by the end of the semester

M.C

Ponca City Senior High School [✓ 10] 11 12

PROGRESS REPORT *GRADE PERIOD* 1 2 [✓ 3] 4 *YEAR* 19 65 19 66

NAME [redacted], Jackie HOMEROOM R. Eskew

SUBJECT Geometry TEACHER Mrs Christ[illegible]

GRADE FOR QUALITY OF ACHIEVEMENT - - SCHOLARSHIP - - SKILL C

Grading system for solid credit subjects: A Excellent B Above average C Averag

D Below average F Failing I Incomplete (to be changed)

Grading system for study halls: S Satisfactory U Unsatisfactory

OTHER QUALITIES AND FACTORS TO BE CONSIDERED

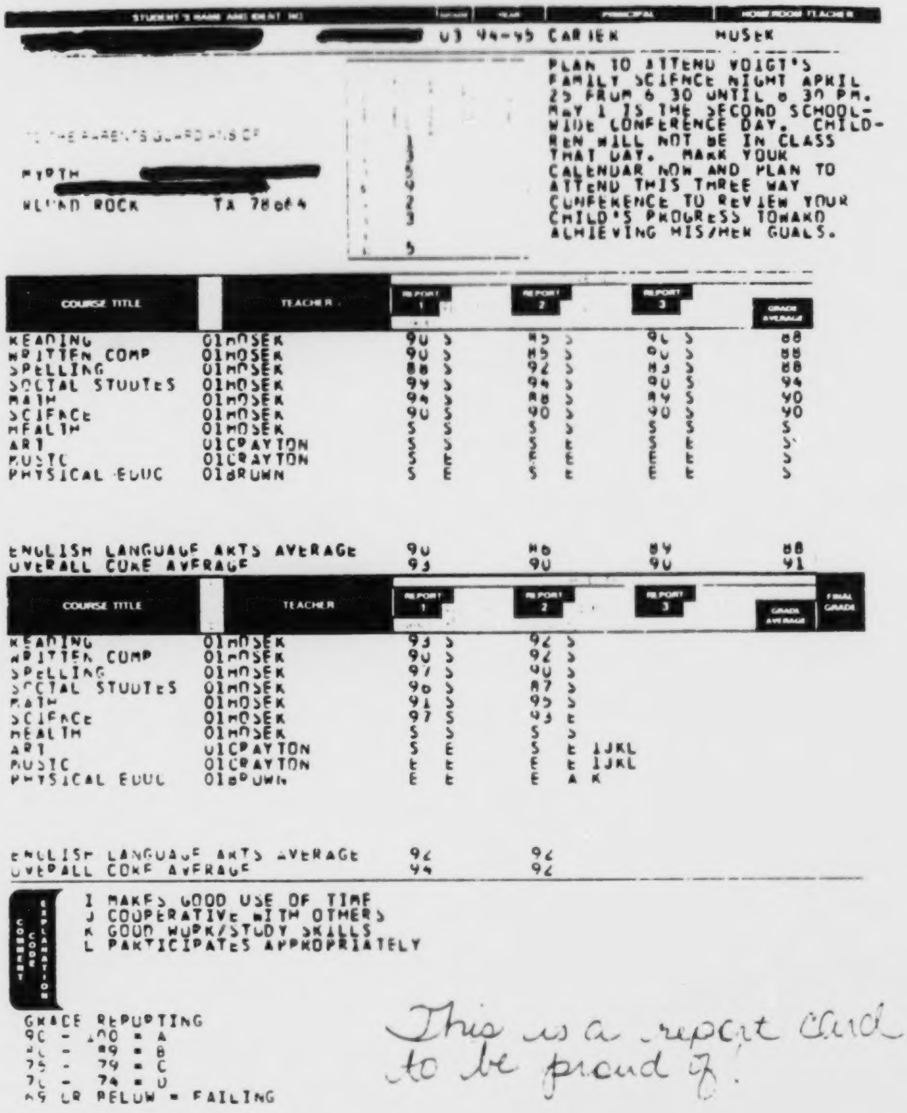
	Good	Average	Poor
1 Work habits - - effort put forth		✓	
2. Citizenship - - attitude - cooperation		✓	

3. Attendance Days absent 0 Times tardy 0

ROBERT E. FORD, *Principal* RO 5-550.

BEST COPY AVAILABLE

Grade 3 Report Card, 1994-95: Myrth



STUDENT'S NAME AND IDENT NO	GRADE	YEAR	PRINCIPAL	HOMEROOM TEACHER
[redacted] [redacted]	03	94-95	CARTER	MOSEK

TO THE PARENTS/GUARDIANS OF

MYRTH [redacted]
[redacted]
ROUND ROCK TX 78664

3
3
5
4
2
3
5

PLAN TO ATTEND VOIGT'S FAMILY SCIENCE NIGHT APRIL 25 FROM 6 30 UNTIL 8 30 PM. MAY 1 IS THE SECOND SCHOOL-WIDE CONFERENCE DAY. CHILDREN WILL NOT BE IN CLASS THAT DAY. MARK YOUR CALENDAR NOW AND PLAN TO ATTEND THIS THREE WAY CONFERENCE TO REVIEW YOUR CHILD'S PROGRESS TOWARD ACHIEVING HIS/HER GOALS.

COURSE TITLE	TEACHER	REPORT 1		REPORT 2		REPORT 3		GRADE AVERAGE
READING	01MOSEK	90	S	85	S	92	S	88
WRITTEN COMP	01MOSEK	90	S	85	S	90	S	88
SPELLING	01MOSEK	88	S	92	S	83	S	88
SOCIAL STUDIES	01MOSEK	99	S	94	S	90	S	94
MATH	01MOSEK	94	S	88	S	89	S	90
SCIENCE	01MOSEK	90	S	90	S	90	S	90
HEALTH	01MOSEK	S	S	S	S	S	S	S
ART	01CRAYTON	S	S	S	E	S	E	S
MUSIC	01CRAYTON	S	E	E	E	E	E	S
PHYSICAL EDUC	01BROWN	S	E	S	E	E	E	S
ENGLISH LANGUAGE ARTS AVERAGE		90		86		89		88
OVERALL CORE AVERAGE		93		90		90		91

COURSE TITLE	TEACHER	REPORT 1		REPORT 2			REPORT 3	GRADE AVERAGE	FINAL GRADE
READING	01MOSEK	93	S	92	S				
WRITTEN COMP	01MOSEK	90	S	92	S				
SPELLING	01MOSEK	97	S	90	S				
SOCIAL STUDIES	01MOSEK	96	S	87	S				
MATH	01MOSEK	91	S	95	S				
SCIENCE	01MOSEK	97	S	93	E				
HEALTH	01MOSEK	S	S	S	S				
ART	01CRAYTON	S	E	S	E	IJKL			
MUSIC	01CRAYTON	E	E	E	E	IJKL			
PHYSICAL EDUC	01BROWN	E	E	E	A	K			
ENGLISH LANGUAGE ARTS AVERAGE		92		92					
OVERALL CORE AVERAGE		94		92					

COMMENT CODE EXPLANATION

I MAKES GOOD USE OF TIME
J COOPERATIVE WITH OTHERS
K GOOD WORK/STUDY SKILLS
L PARTICIPATES APPROPRIATELY

GRADE REPORTING
90 - 100 = A
80 - 89 = B
75 - 79 = C
70 - 74 = D
69 OR BELOW = FAILING

This is a report card to be proud of!

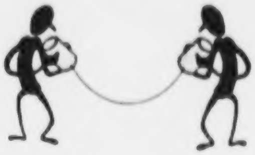
Handout A4.4, H4

BEST COPY AVAILABLE

593

Activity 4.5

Report Card Formats



Purpose:

To examine and critique alternative formats for reporting on student achievement

Uses:

This is an advanced level activity to use in **Chapter 4** to introduce alternative reporting formats. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); (b) understanding of how assessment can be integrated with instruction (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*); and (c) **Activity 4.2**—*Putting Grading and Reporting Issues in Perspective* to set the context for examining report card designs.

Rationale:

Formats for reporting about student achievement must align with current notions of what students should know and be able to do and how to best assist students to achieve them.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A4.5,O1**—*Report Card Formats Purposes*, **A4.5,O2**—*Compare and Contrast*, **A4.5,O3**—*Comparison of Alternative Reporting Formats*
- **Handouts A4.5,H1**—*Round Rock ISD 3-Way*[1]; **A4.5,H2**—*Round Rock ISD Kindergarten*[2]; **A4.5,H3**—*Dos Rios*[3]; **A4.5,H4**—*Shutesbury*[4]; **A4.5,H5**—*Waialae*[5]; **A4.5,H6**—*Crow Island*; **A4.5,H7**—*Wishart*[6] **A4.5,H8**—*Travis Heights*[7]

---

[1] Used with permission of the primary developer, Lynda Carter, Principal, Fern Bluff Elementary School, Round Rock, TX, and published in the *Round Rock Independent School District Parent Conferencing Training Module*, August, 1995. In grades 1-5, biannual conferences supplement traditional 6-weeks report cards.

[2] Used with permission of the primary developers, Lynda Carter and Delores Block, Round Rock ISD, Round Rock TX. The forms replace traditional Kindergarten report cards.

- Additional reporting formats can be found in **Appendix C**—*Articles* (**Article C.2**—*Developmentally Appropriate Report Cards…*) and **Appendix A**—*Sampler* (**Sample A.42**—*Science Portfolio* or **A.8**—*Primary Math Portfolio*)

Time Required:

40-75 minutes

Facilitator Notes:

1. (2 minutes) Use **Overhead A4.5,O1**—*Report Card Formats Purposes*, as needed to introduce the purpose of the activity.
2. (5-10 minutes) Review and make sure participants agree with the idea that grades should:
   - encourage learning and support student success
   - avoid undesirable side effects
   - be meaningful/understandable to the recipient of the information (students and teachers)
   - have the primary purpose of accurately communicating about student achievement

   Participants will have difficulty discussing the advantages of different report card formats if they don't agree on the purposes for grading and reporting.
3. (20-30 minutes) Use **Overhead A4.5,O2**—*Compare and Contrast* to guide small groups of 3-5 participants in comparing and contrasting the report card formats chosen from the handout list. Any single group is likely to have time to review only two or three formats; they can choose or be assigned samples.
4. (10-30 minutes) Have groups discuss advantages and disadvantages of alternative reporting formats. Use **Overhead A4.5,O3**—*Comparison of Alternative Reporting Formats* to record participant responses and summarize reasons for using alternative formats.

🗀 Reference Box

For more information on the report card formats used in this activity:

Crow Island School
112 Willow Road
Winnetka, IL 60093

Dos Rios Elementary School
2201 34th St.
Evans, CO 80620

Round Rock ISD
Fern Bluff Elementary
17815 Park Valley Drive
Round Rock, TX 78781

Shutesbury Elementary School
West Pelham Road
Shutesbury, MA 10172

Travis Heights Elementary School
2010 Alameda Drive
Austin, TX 78704

Waialae Elementary School
1045 19th Avenue
Honolulu, HI 96816

---

[3] Used with permission of Helen Elliott, Principal, Dos Rios Elementary Schools, Evans, CO.

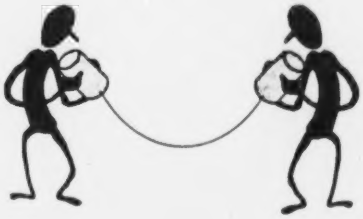
[4] Used with permission of Shutesbury Elementary School, Shutesbury, MA.

[5] Used with permission of Waialae Elementary School, Honolulu, HI. The forms are works in progress.

[6] Used with permission of Wishart Elementary School, School District #62, Sooke.

[7] Used with permission of Carole Garcia, Assistant Principal, Travis Heights Elementary School, Austin, TX.

Report Card Formats



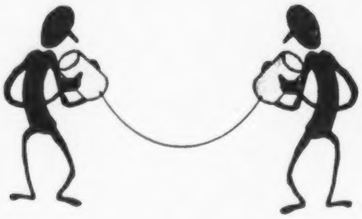
Purpose:

To examine and critique alternative formats for reporting on student achievement

Overhead A4.5,O1

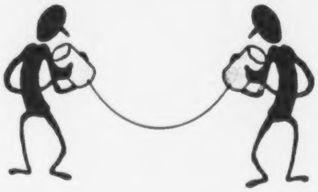
596

Compare and Contrast



- How are they alike?
- How are they different?
- What does each report card convey about the student's knowledge, skills, and understanding?
- Who is the target audience? Would the target audience understand the message sent?
- Which would best encourage learning and support student success?
- Which would best avoid undesirable side effects?

Comparison of Alternative Reporting Formats

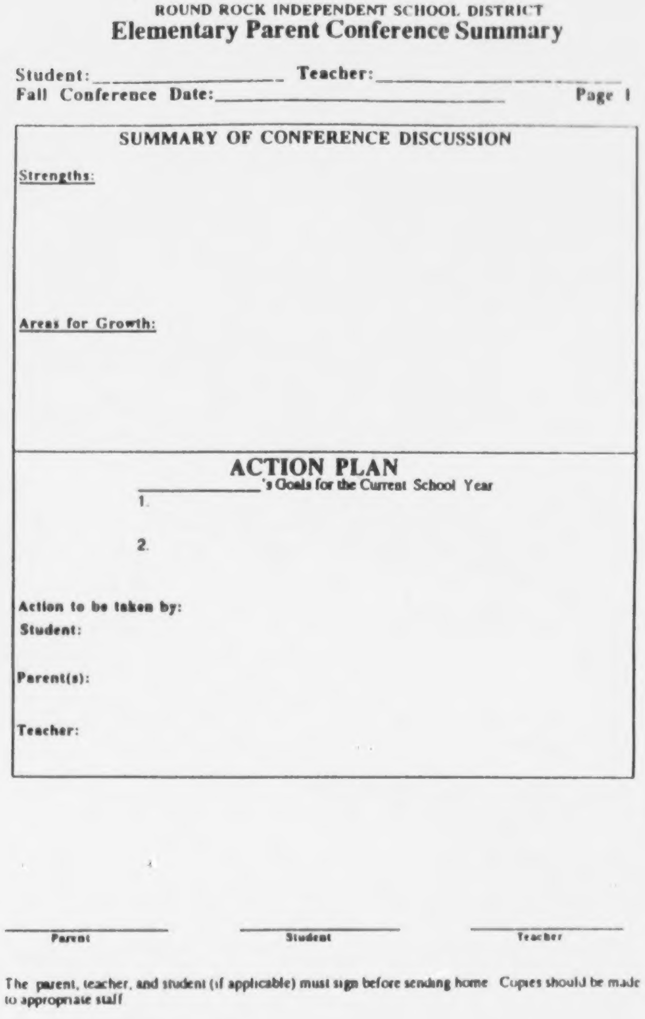


Sample #	Advantages	Disadvantages

Overhead A4.5,O3

Round Rock ISD 3-Way Conference Reports
Elementary Parent Conference Summary Forms

Handout A4.5,H1, p. 1



ROUND ROCK INDEPENDENT SCHOOL DISTRICT
Elementary Parent Conference Summary

Student:__________ Teacher:__________
Fall Conference Date:__________ Page 1

SUMMARY OF CONFERENCE DISCUSSION

Strengths:

Areas for Growth:

ACTION PLAN

________'s Goals for the Current School Year

1.

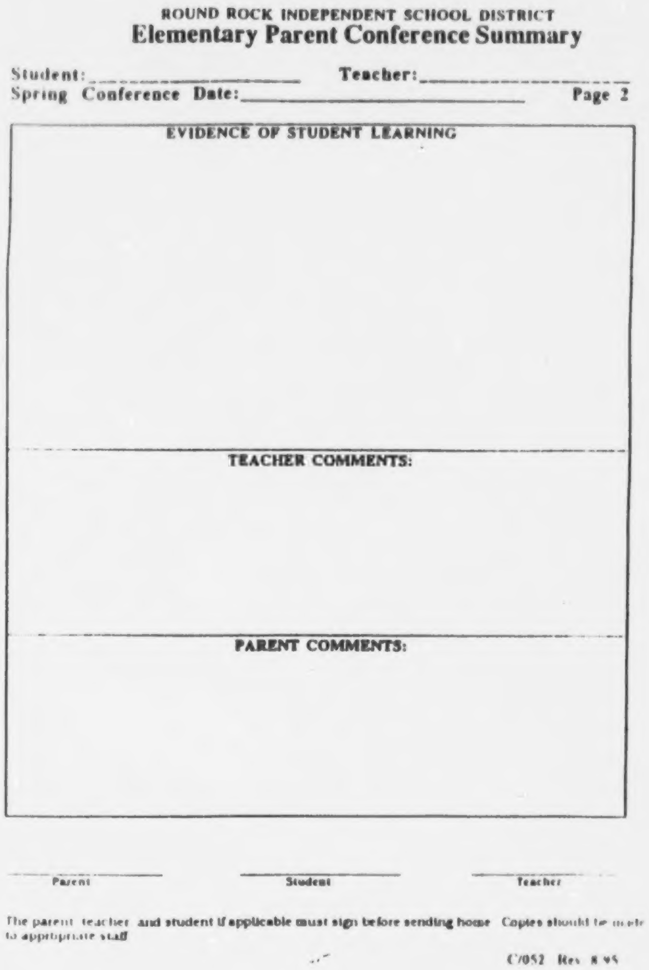
2.

Action to be taken by:
Student:
Parent(s):
Teacher:

Parent Student Teacher

The parent, teacher, and student (if applicable) must sign before sending home. Copies should be made to appropriate staff

C/050 Rev. 8.95



ROUND ROCK INDEPENDENT SCHOOL DISTRICT
Elementary Parent Conference Summary

Student:__________ Teacher:__________
Spring Conference Date:__________ Page 2

EVIDENCE OF STUDENT LEARNING

TEACHER COMMENTS:

PARENT COMMENTS:

Parent Student Teacher

The parent, teacher, and student (if applicable) must sign before sending home. Copies should be made to appropriate staff

C/052 Rev. 8.95

BEST COPY AVAILABLE

Shared Goal Setting

1. The teacher and students prepare for the upcoming conference. This includes building rapport with parents through informal communication (notes home, phone calls, visits at school), sharing evidence of learning through weekly work sent home, newsletters, and sharing with them student reflections about their work.

2. The teacher helps parents prepare to attend the conference by sending home a letter that outlines the conference format and a Parent Information Sheet* seeking parent input regarding their child's perceived strengths and areas for future growth.

3. Before each conference the student and teacher spend time reviewing the child's collection of work to be shared at the conference as evidence of student progress over time.

4. The parent(s), student, and teacher then meet in the classroom conference area. A discussion is conducted that involves what has been learned, and what areas the student needs to improve upon. Proposed learning goals for the year are determined through collaboration.

5. Learning goals are agreed upon. In a three-way conference each party--student, parent, and teacher commit to providing specific support in accomplishing goals.

6. The teacher completes a record of the conference and determined goals. Copies: parent, office/counselor

* The Parent Information Sheet may be custom tailored by each campus staff

Round Rock ISD 3-Way Conference Reports
Teacher Information Sheet:
Shared Goal Setting

Handout A4.5,H1, p. 2

601

Three-Way Conference Guide for Students

Name: ____________ Date: ____________

Two things I do well.......
Two things I need to improve........
Things I want to show.......
My goals for the first semester are.......

Modified chart from Together is Better by Cameron, Davies, Politano, and Gregory

Round Rock ISD 3-Way Conference Reports
Guide Sheet for Students

602

Handout A4.5,H1, p. 3

A Conference Guide for Teachers

Student's Name: ____________________ **Date:** ________________________

Areas of Strength:	**Areas for Future Growth:**
Notes for conference:	**Additional Notes:**
Action Plan: Goals for First Semester:	Evidence of Learning:
Goals for Second Semester:	Evidence of Learning:

Round Rock ISD 3-Way Conference Reports
Conference Guide for Teachers

BEST COPY AVAILABLE

Handout A4.5, H1, p. 4

Name ______________________________ Date ______________

Child's name ________________________ Grade ______________

Parent Interview

1 How do you think your child is doing as a reader/writer? Why? (If a young child: What signs have you seen that your child is ready to learn to read/write?

____

____

____

____

2 What would you like your child to do as a reader/writer that he or she isn't doing now?

____

____

____

____

3 Do you ever notice your child reading/writing at home? Tell me about it.

____

____

____

____

4 What do you think your child's attitude is toward reading/writing? What do you think has helped to create this attitude?

____

____

____

____

5 What sorts of questions about your role in helping your child become a better reader/writer might you like to ask me?

____

____

6 Since I like to help the children read and write about things they are interested in, it helps me to know each individual child's interests. What kinds of things does your child like to do in his or her free time?

____

____

____

____

7 Is there anything about the child's medical history that might affect his or her reading/writing? Is there anything else that might affect his or her reading/writing?

____

____

____

____

8 Is there anything else that you think would be helpful for me to know in teaching your child?

____

____

Round Rock ISD 3-Way Conference Reports
Parent Interview Form

604

BEST COPY AVAILABLE

Round Rock ISD
Kindergarten Learning Profile
Page 1

ROUND ROCK INDEPENDENT SCHOOL DISTRICT
KINDERGARTEN LEARNING PROFILE

A learning profile reflects growth of the kindergarten child in different areas of development. This profile corresponds to the Texas Kindergarten Essential Elements and the objectives of Round Rock Independent School District and offers opportunity for growth beyond kindergarten expectations. It is designed to communicate progress in a developmentally appropriate, interdisciplinary, and interactive curriculum.

LABEL

1995-1996 PARENT CONFERENCE DATES

OCTOBER 3, 1995
APRIL 24, 1996

PARENT SIGNATURE		TEACHER CONFERENCE REQUEST	
Please sign and return	Conference Requested ☑		Reasons
2. ____________	☐		
3. ____________	☐	☐ 2.	____________
4. ____________	☐	☐ 3.	____________
5. ____________	☐	☐ 4.	____________

C 6/95 003

606

605

BEST COPY AVAILABLE

Round Rock ISD
Kindergarten Learning Profile
Page 2

LISTENING	●Listens and attends sporadically	●Listens and attends for short periods of time ●Inconsistently responds to verbal commands and instruction	●Listens and attends for longer periods of time during individual and group interactions ●Responds to speaker	●Listens and attends consistently to respond to ideas, feelings, information, and experiences	●Listens and attends with little or no prompting to recall many details	●Listens critically and attends to make connections between topics and situations
SPEAKING	●Communicates nonverbally (gestures, eye-contact, nodding, facial expression) ●Expresses with limited verbal skills	●Uses appropriate language to express basic physical, emotional, and social needs ●Uses appropriate voice tone to express emotions	●Verbally contributes information/ideas with prompting, in small group and one-to-one ●Participates in choral speaking opportunities and informal creative dramatics	●Expresses ideas, feelings, information, and experiences in group settings and through creative dramatics, sharing, or show and tell ●Uses concept vocabulary appropriately	●Expresses ideas and feelings with fluency ●Asks questions and shares ideas relevant to topic ●Uses voice tone to role-play	●Combines information to draw conclusions and communicate ideas with others
WRITING	●Expresses ideas through drawing and painting ●Uses writing materials ●Scribbles randomly	●Begins to understand that print conveys a message ●Shows appropriate use of writing materials ●Scribbles with meaning using letter like shapes or random letters ●Dictates words, phrases, or sentences	●Associates print with spoken language ●Shows awareness of letter sound associations ●Shows awareness of left to right ●Copies environmental print ●Writes some high frequency words like mom, dad	●Writes words or phrases using letter(s) to represent sound(s) (some reversals) ●Often uses two or three letters to represent longer words ●Writes left to right, top to bottom	●Attempts to record own ideas, mostly independently ●Writes simple sentences using temporary spelling (some reversals) ●Explores capitalization and punctuation ●Spaces between words	●Extends and elaborates ideas ●Applies capitalization, punctuation, and spacing ●Demonstrates readiness for more direct spelling instruction ●Arranges and spaces text without difficulty ●Begins to revise and edit own work

Handout A4.5, H2, p. 2

BEST COPY AVAILABLE

Round Rock ISD
Kindergarten Learning Profile
Page 3

Handout A4.5, H2, p. 3

PRE-READING/READING

●Develops an awareness of printed materials and pictures	●Displays interest in print ●Spends time looking at books and likes to have stories read	●Retells stories from pictures or media ●Begins to recognize some concepts of print ●Actively participates in shared reading ●Recognizes some names and environmental print in context	●Follows print left-to-right ●Identifies and connects sounds with letters ●Recognizes some high frequency words or names in and out of context(and, the, to, ...) ●Shows signs of becoming a beginning reader	●Participates comfortably in reading activities ●Follows print with 1 to 1 correspondence ●Begins to use phonics, pictures and language structure to read for meaning	●Reads for meaning/ information ●Uses reading strategies automatically ●Is comfortable with a variety of literature

INTEGRATED CONTENT

●Randomly observes and explores concrete objects.	●Expresses interest consistently in concepts. Observes and explores concrete objects.	●With guidance, demonstrates understanding of concepts by manipulating concrete objects with a purpose.	●Independently demonstrates understanding of concepts by manipulating concrete objects. Gives clear explanations to show understanding.	●Connects concrete to symbolic. Applies knowledge to solve problems.	●Independently works at symbolic level to interpret and extend concepts.

MATH CONCEPTS

Number Concepts	Comparison	Measurement
counting	colors	calendar
one to one	shapes	graphs
sets	patterns	money
ordinal numbers	position	time
fractions	sort/classify	size

SOCIAL STUDIES CONCEPTS

Self
Family
Community/School
Celebration/Cultural Events
Geography

SCIENCE CONCEPTS

Habitats of Animals/Plants in our Community
Measuring
Responding to the Environment
Safety
Seasons and their Effects

BEST COPY AVAILABLE

Round Rock ISD
Kindergarten Learning Profile
Page 4

MUSIC

- Shows interest/participates in chants, nursery rhymes, folk songs and singing games
- Listens to a variety of musical selections/sounds
- Responds to music with movement and creative expressions

- Accurately recognizes singing vs. speaking voice
- Listens to music comparatives
  - fast/slow
  - high/low
  - loud/soft
  - staccato/legato
- Moves in response to music with a sense of beat

- With guidance, sings simple echo patterns in small groups
- With guidance, demonstrates understanding of music comparatives
- With guidance, demonstrates steady beat in response to music

- Independently sings simple echo patterns
- Distinguishes basic instrument sounds
- Independently demonstrates steady beat

- Independently sings a familiar song
- Independently uses/demonstrates music comparatives:
  - fast/slow
  - staccato/legato
  - loud/soft
  - upward/downward
  - high/low
  - long/short
- Distinguishes between steady beat and rhythm

- Independently sings in tune
- Differentiates instrumental/choral/solo selections
- Creates rhythmic patterns

In addition to the singing concepts, comparatives and listening skills, rhythm, movement, and expression reflected on this continuum, the music curriculum includes multicultural songs and dances, listening activities, singing games, creative dramatics, instrumental experiences, cross curriculum reinforcement, seasonal and patriotic songs, instruments/families of the orchestra, performance experiences, American folk songs and dances and lessons that introduce melodic and rhythmic reading

VISUAL ARTS

- Explores and observes during the art production process
- Marks with media are uncontrolled and undirected

- Discovers and becomes aware of art elements
- Makes controlled lines with uncontrolled motions
- Unrecognizable images are created

- Recognizes, examines and responds utilizing art elements during the production process
- Scribbling more controlled
- Images have some recognizable form

- Identifies art elements
- Demonstrates skills and experiments with media
- Images are recognizable

- Applies knowledge of art elements and media skills during production activities
- Images contain significant details
- Use of background and horizon line can be employed

- Utilizes individual thoughts and ideas into art production process
- Composition of images becomes more complex with inclusion of background and use of the whole picture plan

VISUAL ARTS ELEMENTS

Line
Shape
Color
Texture

PHYSICAL EDUCATION

- Travels in different ways without bumping or falling
- Identifies selected body parts
- Participates daily in vigorous physical activity
- Recognizes safe and effective learning behaviors

- Walks and runs with balance and control
- Changes directions quickly on cue
- Demonstrates shapes, patterns, relationships
- Identifies effects of physical activity
- Participates appropriately and safely; alone and with others

- Tosses ball and catches after one bounce
- Kicks a stationary ball with control
- Rolls body sideways
- Relates skill development to practice
- Identifies and demonstrates safe and effective learning behaviors

- Throws overhand and underhand
- Repeatedly jumps a rope turned by others
- Sustains moderate physical activity
- Relates physical activity to well-being
- Consistently uses safe and effective learning behaviors

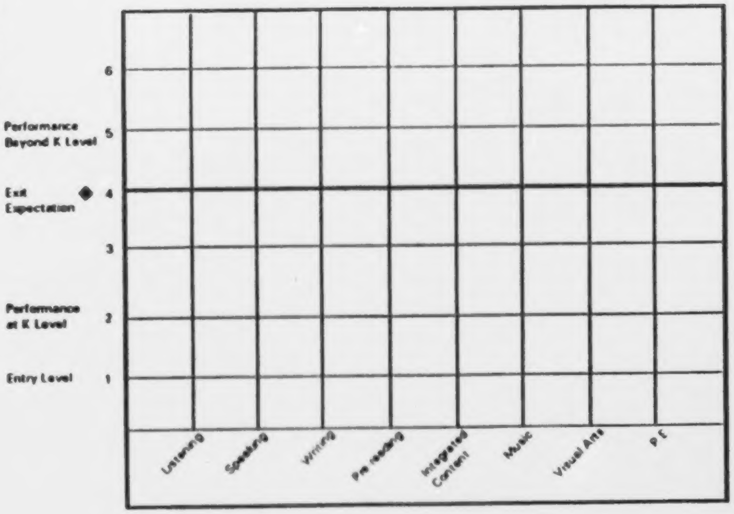
- Travels backward without falling
- Strikes stationary objects with control
- Kicks moving objects with control
- Coordinates movement to rhythms and music
- Demonstrates appropriate, safe, and social behaviors

- Dribbles (hand/foot) with control
- Catches using properly positioned hands
- Repeatedly jumps a self-turned rope
- Supports own body weight while hanging and climbing
- Cooperates and shares readily

BEST COPY AVAILABLE

Round Rock ISD
Kindergarten Learning Profile
Page 5

KINDERGARTEN LEARNING PROFILE SUMMARY



Blue	1st Six Wks (29)	9-22-95	Green	4th Six Wks (29)	2-15-96
Red	2nd Six Wks (27)	11-3-95	Purple	5th Six Wks (29)	4-10-96
Black	3rd Six Wks (26)	12-14-95	Orange	6th Six Wks (30)	5-23-96

LABEL

ATTENDANCE

	1	2	3	4	5	6	TOTAL
Days Absent							
Days Present							
Tardies							

To Grade ____________ 19 ____________

☐ PLACED ☐ PROMOTED ☐ RETAINED

BEST COPY AVAILABLE

READING INDICATORS - Primary

Emergent Reader

- •Has print awareness
- •Knows letter names and sound-symbol correspondence (single consonant, blends, and consonant digraphs
- •Matches spoken words to print
- •Understands the concept of a word
- •Combines letter sounds and uses picture clues to read words
- •Chooses material at independent reading level
- •Has a store of sight words in reading and writing
- •Uses a variety of strategies:
  - Rereads when it doesn't make sense
  - Self-corrects errors
  - Relies on visual cues
- •Comprehends what has been read; can retell a story

Developing Reader

- •Uses word attack strategies and/or cueing systems when reading for meaning (phonics, semantics, syntax, and picture clues)
- •Chooses material at independent reading level
- •Uses a store of sight words in reading and writing
- •Uses a variety of strategies to interact (construct meaning) with both fiction and non-fiction text (strategies i.e. semantic web, K.W.L., DR-TA, story maps, etc.)
- •Comprehends what has been read; literal and inferential
- •Reads with fluency and expression
- •Responds effectively to text

Fluent Reader

- •Chooses to read a variety of materials for a variety of purposes
- •Has the ability to construct meaning under a variety of different reader, text, and contextual conditions
- •Has the ability to evaluate and integrate information both within a text and from more than one text
- •Knows and uses a variety of strategies for identifying words; e.g., predictions, context clues, phonics, and structural analysis
- •Knows and uses a variety of strategies to aid comprehension; e.g., summarizing, self-questioning, and predicting.
- •Comprehends at inferential level

WRITING INDICATORS - Primary

Advanced (AD)

☐ The writing meets Essential Level criteria and contains a logical progression of ideas about the topic. Word choice enhances the writing and demonstrates insight, creativity and/or style.

△ The writing contains complete thoughts and consistent use of these essential skills:

•capitalization (beginning of sentences, I, and names of people)
•periods
•varied sentence patterns
Standard spelling of many high frequency words is evident

Essential (ES)

☐ The writing meets Developing Level criteria. The pattern of organization and the transition of ideas flow. The descriptive or supporting details are adequate and are appropriate to the purpose of the writing.

△ The writing contains complete thoughts and consistent use of these essential skills:
•capitalization (beginning of sentences, I, and names of people)
•periods
•varied sentence patterns
Standard spelling of many high frequency words is evident.

In addition to close approximations to standard spelling throughout, the writing contains standard spelling of most high frequency words.

Developing (DV)

☐ The writing is a series of related ideas about the task. All of the ideas can be readily understood and convey a sense of completeness.

△ The writing contains complete thoughts and emergent use of the essential skills:
•capitalization (beginning of sentences, I, and names of people)
•periods
•varied sentence patterns

In addition to close approximations to standard spelling throughout, the writing contains standard spelling of some high frequency words.

In Progress (IP)

☐ The writing contains ideas which cannot be readily understood (unfocused) and/or contains an insufficient number of related thoughts to comprise a message.

△ The writing contains incomplete thoughts and/or sound symbol association is in progress.

☐ Primary Trait (PT)
△ Secondary Trait (ST)

INDIVIDUAL
PRIMARY PROGRESS REPORT

DOS RIOS ELEMENTARY SCHOOL
Weld County School District Six

2201 34th Street
Evans, CO 80620
(303) 330-5957

(place label here)

Dos Rios Elementary School
Primary Progress Report, Page 1

BEST COPY AVAILABLE

Dos Rios Elementary School
Primary Progress Report, Page 2

READING

Emergent		
Developing		
Fluent		

(Check where the student is currently performing in reading skills)

WRITING (criteria on back)

DEVELOPMENT/ORGANIZATION OF IDEAS (PT)		
Narrative		
GRAMMAR, MECHANICS, SPELLING, STRUCTURE (ST)		
Narrative		

KEY

4 = Exceeds expectations for age level
3 = Effectively demonstrates
2 = Demonstrates some of the time
1 = Needs more time and practice
NA = Not assessed at this time

PERSONAL-SOCIAL DEVELOPMENT

Respects Self		
Respects Others		
Respects Property		
Use of Problem Solving Strategies		
Applies Cooperative Learning Strategies		

ATTENDANCE	1	2	3	TOTAL
Days Present				
Days Absent				
TOTAL				

MATH

Problem Solving		
Mathematical Connections		
Reasoning		
Estimation		
QUANTITY & NUMBER		
Number Sense		
Concepts of Whole Numbers		
Whole Number Computation		
Rational Numbers (fractions & decimals)		
SHAPES/GEOMETRY/MEASUREMENT		
Geometry		
Measurement (standard & metric)		
Time & Money		
PATTERNS/ALGEBRA		
Patterns & Relationships		
Algebra		
GRAPHS/DATA/STATISTICS		
Statistics & Probability		

Placement for next year ______________________

SCIENCE/HEALTH ☐ Change ☐ Patterns		
Uses an Inquiry Approach		
Uses Process skills		
Understands Science Concepts		
SOCIAL STUDIES		
Understands Concepts & Application of		
Geography		
Political/Civic Responsibility		
History/Future Perspective		
Cultural Diversity		
Economics		
INTEGRATED STUDIES		
Applies reading, writing, math, social studies, science, & health to communicate about		

COMMENTS

____
____
____
____

BEST COPY AVAILABLE

READING INDICATORS - Intermediate

Emergent Reader

- Has print awareness
- Knows letter names and sound-symbol correspondence (single consonants, blends, and consonant digraphs
- Matches spoken words to print
- Understands the concept of a word
- Combines letter sounds and uses picture clues to read words
- Chooses material at independent reading level
- Has a store of sight words in reading and writing
- Uses a variety of strategies:
  - Rereads when it doesn't make sense
  - Self-corrects errors
  - Relies on visual cues
- Comprehends what has been read; can retell a story

Developing Reader

- Uses word attack strategies and/or cueing systems when reading for meaning (phonics, semantics, syntax, and picture clues)
- Chooses material at independent reading level
- Uses a store of sight words in reading and writing
- Uses a variety of strategies to interact (construct meaning) with both fiction and non-fiction text (strategies i.e. semantic web, K.W.L., DR-TA, story maps, etc.)
- Comprehends what has been read: literal and inferential
- Reads with fluency and expression
- Responds effectively to text

Fluent Reader

- Chooses to read a variety of materials for a variety of purposes
- Has the ability to construct meaning under a variety of different reader, text, and contextual conditions
- Has the ability to evaluate and integrate information both within a text and from more than one text
- Knows and uses a variety of strategies for identifying words; e.g., predictions, context clues, phonics, and structural analysis
- Knows and uses a variety of strategies to aid comprehension; e.g., summarizing, self-questioning, and predicting.
- Comprehends at inferential level

WRITING INDICATORS - INTERMEDIATE

Advanced (AD)

☐ The writing meets Proficient Level criteria and contains examples of one or more of the following: insight, creativity, fluency, critical thinking or style.

△ The writing exhibits control of writing conventions beyond those at the Essential Level. Sentence length and complexity complement the writing.

Proficient (PR)

☐ The writing meets Essential Level criteria and contains a logical progression of ideas about the topic. The pattern of organization and the transition of ideas flow. Word choice enhances the writing.

△ The writing contains consistent application of Essential Level skills and standard spelling of high frequency words relative to length and complexity of the writing. Varied sentence patterns are present.

Essential (ES)

☐ The writing is a series of related ideas conveying the writer's interpretation of the task within the mode of type of writing discourse. The pattern of organization and the descriptive or supporting details are adequate and are appropriate to the mode of discourse.

△ The writing contains first draft application of these Essential Level skills:
- complete sentences
- capitalization (proper nouns, sentence beginnings, "I")
- end punctuation

Standard spelling of high frequency words is evident.

In Progress (IP)

☐ The writing is an underdeveloped and/or unfocused message.

△ The writing contains a pattern of errors in one or more of these Essential level skills:
- complete sentences
- capitalization (proper nouns, sentence beginnings, "I")
- end punctuation

A pattern of errors in spelling high frequency words is evident.

☐ Primary Trait (PT)
△ Secondary Trait (ST)

INDIVIDUAL
INTERMEDIATE PROGRESS REPORT

DOS RIOS ELEMENTARY SCHOOL
Weld County School District Six

2201 34th Street
Evans, CO 80620
(303) 330-5957

(place label here)

Dos Rios Elementary School
Intermediate Progress Report
Page 1

BEST COPY AVAILABLE

Dos Rios Elementary School
Intermediate Progress Report
Page 2

READING

Emergent		
Developing		
Fluent		

(Check where the student is currently performing in reading skills)

WRITING

DEVELOPMENT/ORGANIZATION OF IDEAS (PT)		
Narrative		
Persuasive		
GRAMMAR, MECHANICS, SPELLING, STRUCTURE (ST)		
Narrative		
Persuasive		

District Assessments:

	Score.	Date.
Mathematics	______	______
Reading:		
Story Selection	______	______
Information	______	______
Writing:		
Narrative		
-Development of ideas	______	______
-Grammar, Spelling	______	______
Persuasive		
-Grammar, Spelling	______	______

A = Advanced
P = Proficient (Essential or above
E = Essential is required to meet
I = In Progress District Standards)
NA = Not assessed at this time

KEY
4 = exceeds expectations for age level
3 = Effectively demonstrates
2 = Demonstrates some of the time
1 = Needs more time and practice
NA = Not assessed at this time

Placement for next year: ______________________

MATH

Problem Solving		
Mathematical Connections		
Reasoning		
Estimation		
QUANTITY & NUMBER		
Number Sense		
Concepts of Whole Numbers		
Whole Number Computation		
Rational Numbers (fractions & decimals)		
SHAPES/GEOMETRY/MEASUREMENT		
Geometry		
Measurement (standard & metric)		
Time & Money		
PATTERNS/ALGEBRA		
Patterns & Relationships		
Algebra		
GRAPHS/DATA/STATISTICS		
Statistics & Probability		

ATTENDANCE	1	2	3	TOTAL
Days Present				
Days Absent				
TOTAL				

SCIENCE/HEALTH	☐ Energy	☐ Systems
Uses an Inquiry Approach		
Uses Process Skills		
Understands Science Concepts		
SOCIAL STUDIES		
Understands Concepts & Application of:		
Geography		
Political/Civic Responsibility		
History/Future Perspective		
Cultural Diversity		
Economics		
INTEGRATED STUDIES		
Applies reading, writing, math, social studies, science, & health to communicate about:		

PERSONAL-SOCIAL DEVELOPMENT

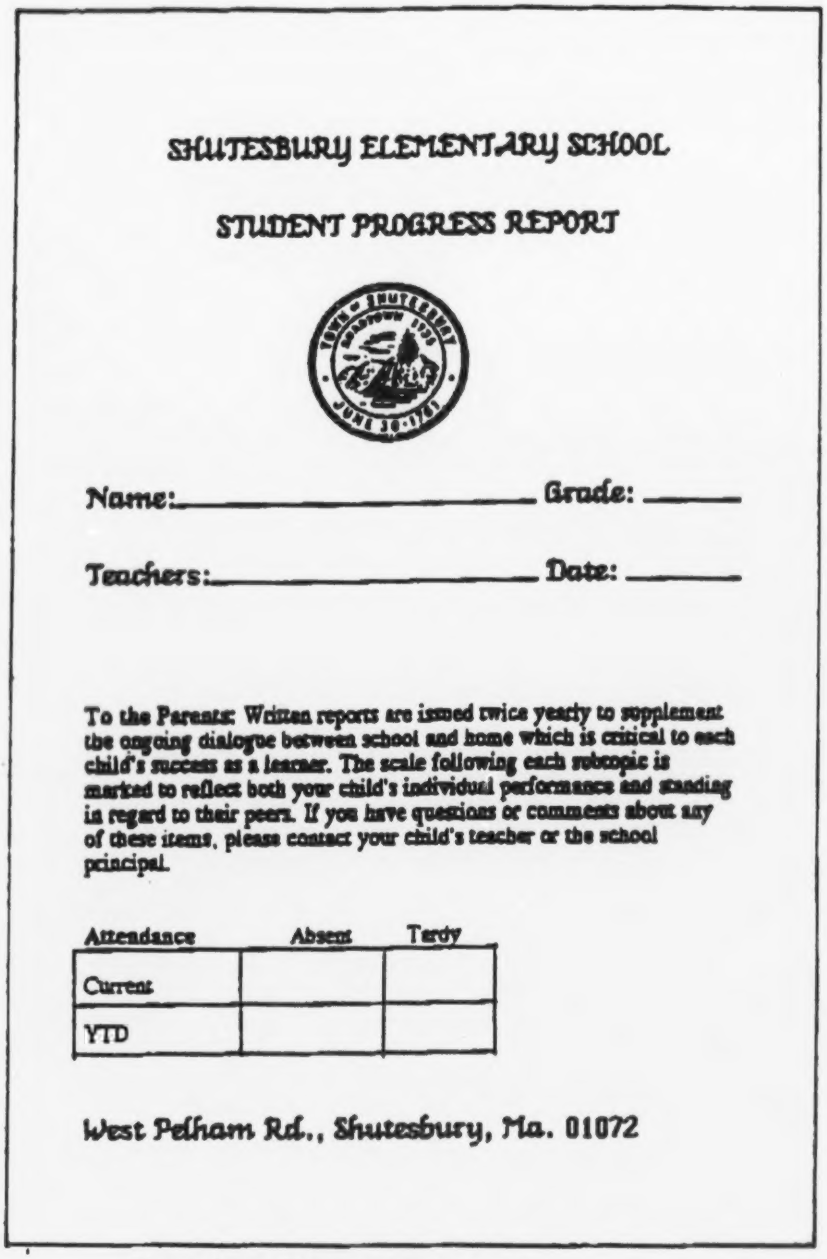
Respects Self		
Respects Others		
Respects Property		
Use of Problem Solving Strategies		
Applies Cooperative Learning Strategies		

621

BEST COPY AVAILABLE

622

Shutesbury Elementary Progress Reports, Grades 3-6



SHUTESBURY ELEMENTARY SCHOOL

STUDENT PROGRESS REPORT

JUNE 30·1761

Name:____________________ **Grade:** ______

Teachers:____________________ **Date:** ______

To the Parents: Written reports are issued twice yearly to supplement the ongoing dialogue between school and home which is critical to each child's success as a learner. The scale following each subtopic is marked to reflect both your child's individual performance and standing in regard to their peers. If you have questions or comments about any of these items, please contact your child's teacher or the school principal.

Attendance	Absent	Tardy
Current		
YTD		

West Pelham Rd., Shutesbury, Ma. 01072

Shutesbury Elementary School
Student Progress Report
Page 1

BEST COPY AVAILABLE

Shutesbury Elementary School
Student Progress Report
Page 2

LANGUAGE ARTS

READING	Area of Concern	Area of Excellence
Level of Confidence		
Uses Context		
Uses Letter/Sound Cues		
Sight Vocabulary		
Reads for Meaning		
Enjoys Books		
Reads Independently		
LISTENING		
Attentively		
With Understanding		
WRITTEN EXPRESSION		
Clarity		
Organization		
Language Content		
Functional Spelling		
Mechanics		
Neatness		
ORAL EXPRESSION		

MATH

Units/Skills Covered: ____________________

____

	Area of Concern	Area of Excellence
Computes Accurately		
Understands Concepts		
Problem Solving		
Level of Confidence		
Keeps Up With Class and Homework		

SCIENCE/ SOCIAL STUDIES

Units of Study: ____________________

____

	Area of Concern	Area of Excellence
Involvement		
Participation		
Understands Concepts		
Project Work		

BEST COPY AVAILABLE

Shutesbury Elementary School
Student Progress Report
Page 3

PERSONAL DEVELOPMENT

Rarely ______________________ Consistently

Exhibits Self-Confidence ______________________
Respects Ideas and Feelings of Others ______________________
Is Helpful and Considerate ______________________
Demonstrates Self-Control ______________________
Adjusts to New Situations ______________________
Cares for Materials ______________________

WORK HABITS

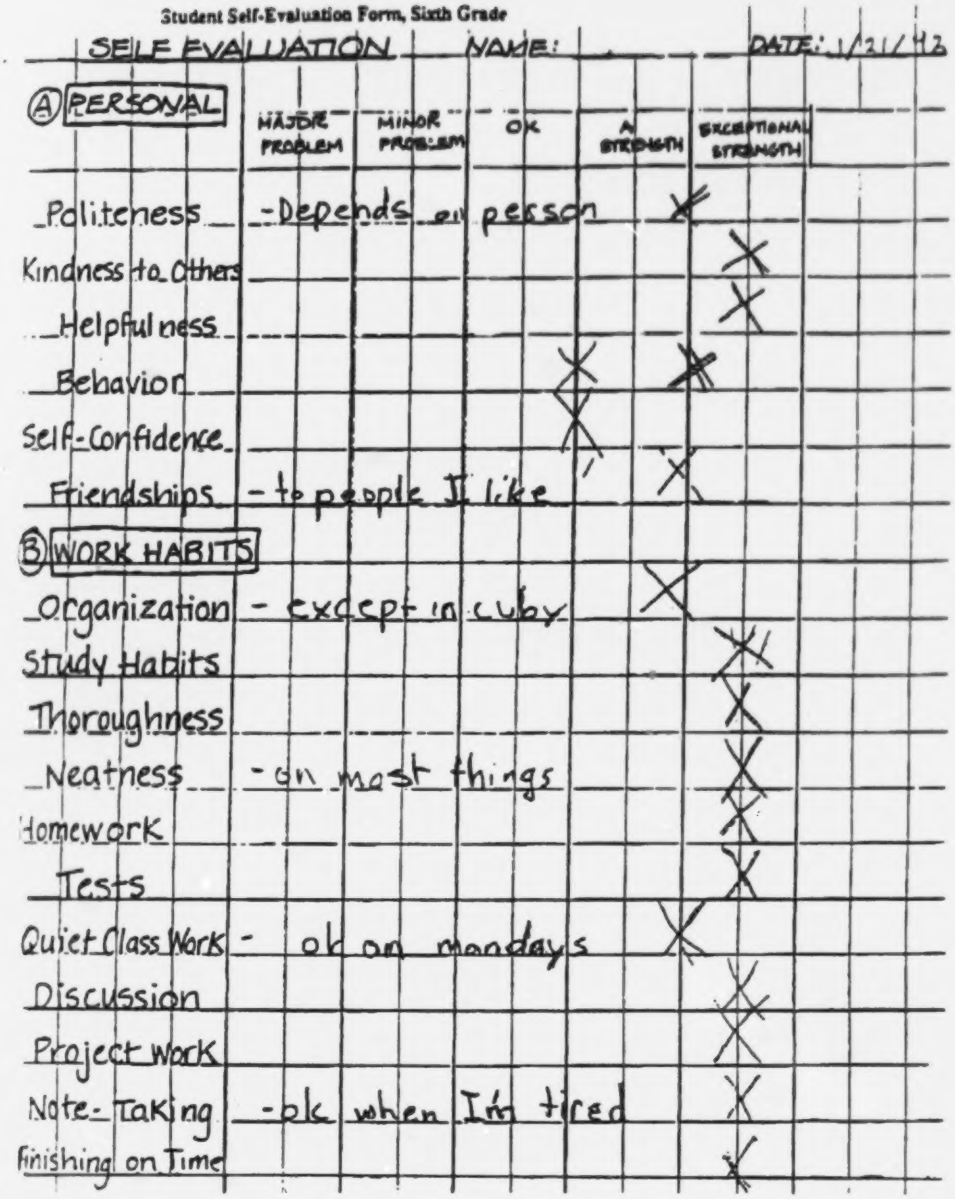
Rarely ______________________ Consistently

Works Well in Groups ______________________
Works Well Alone ______________________
Works Carefully, Thoroughly ______________________
Completes Work On Time ______________________
Presents Work Neatly ______________________
Follows Directions ______________________
Is Organized ______________________

COMMENTS

p. 3

BEST COPY AVAILABLE



Student Self-Evaluation Form, Sixth Grade

SELF EVALUATION NAME: DATE: 1/21/93

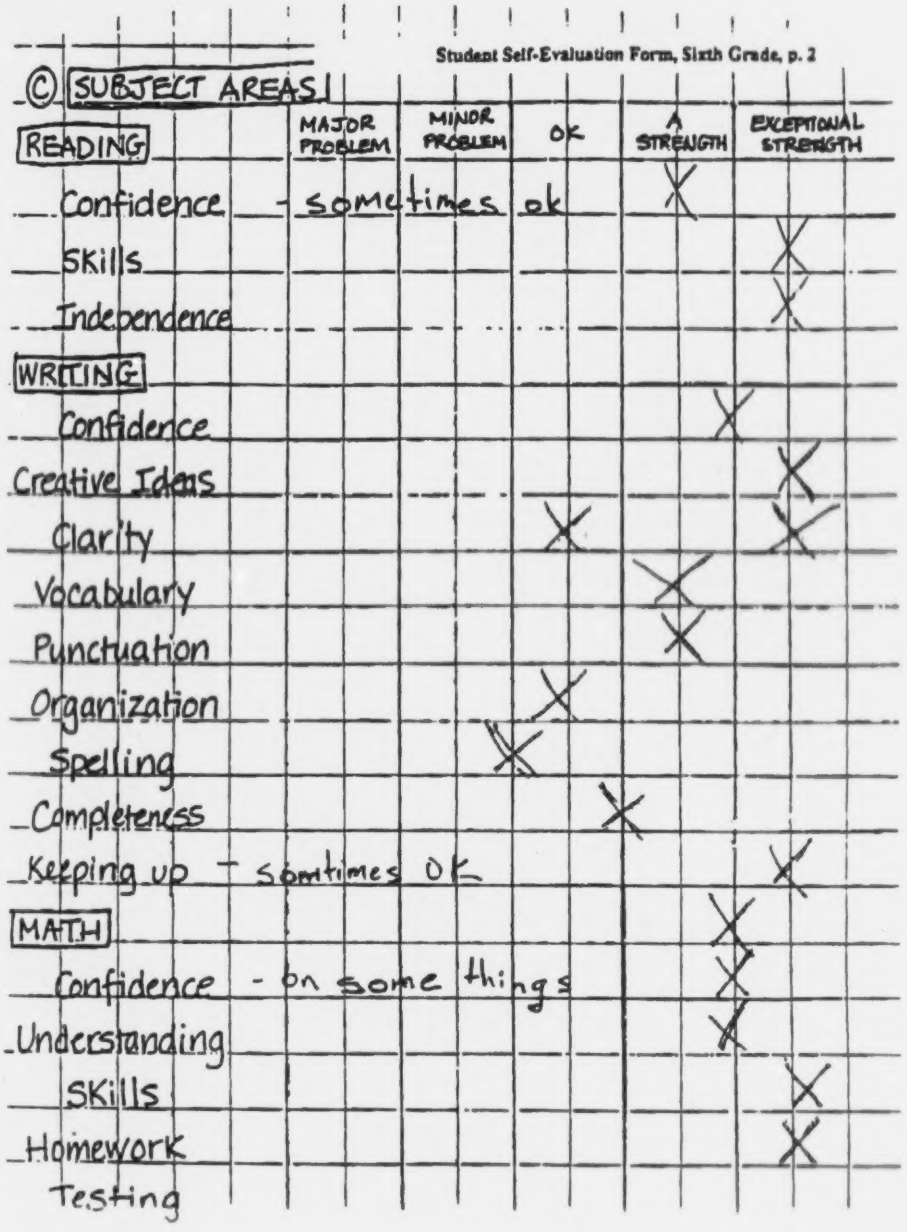
Ⓐ PERSONAL	MAJOR PROBLEM	MINOR PROBLEM	OK	A STRENGTH	EXCEPTIONAL STRENGTH
Politeness	-Depends on person			X	
Kindness to Others					X
Helpfulness					X
Behavior			X	X	
Self-Confidence			X		
Friendships	- to people I like			X	
Ⓑ WORK HABITS					
Organization	- except in cuby			X	
Study Habits					X
Thoroughness					X
Neatness	- on most things				X
Homework					X
Tests					X
Quiet Class Work	- ok on mondays			X	
Discussion					X
Project work					X
Note-Taking	-ok when I'm tired				X
Finishing on Time					X

Shutesbury Elementary School
Student Self Evaluation Form
Sixth Grade
Page 1

BEST COPY AVAILABLE

Handout A4.5, H4, p. 4

628



Student Self-Evaluation Form, Sixth Grade, p. 2

Ⓒ SUBJECT AREAS

	MAJOR PROBLEM	MINOR PROBLEM	OK	A STRENGTH	EXCEPTIONAL STRENGTH
READING					
Confidence - sometimes ok				X	
Skills					X
Independence					X
WRITING					
Confidence				X	
Creative Ideas					X
Clarity			X		X
Vocabulary				X	
Punctuation				X	
Organization			X		
Spelling		X			
Completeness			X		
Keeping up - somtimes OK					X
MATH				X	
Confidence - on some things				X	
Understanding				X	
Skills					X
Homework					X
Testing					

Shutesbury Elementary School
Student Self Evaluation Form
Sixth Grade
Page 2

629

Handout A4.5, H4, p. 5

SCIENCE/SOCIAL STUDIES

		MAJOR PROBLEM	MINOR PROBLEM	OK	A STRENGTH	EXCEPTIONAL STRENGTH
Listening Skills					X	
Notes						X
Written Work						X
Graphic Work					X	
Discussion					X	
Projects					X	
GYM Effort	- Depends on what we do				X	
Skills	- Depends on what we do					X
Behavior					X	
ART Effort	- Depends on what we do					X
Skills	- Depends on what we do				X	
Behavior					X	
MUSIC Effort	- Depends on what we do			X		
Skills	- Depends				X	
Behavior	- Depends			X		
SPORTS				X		

Shutesbury Elementary School
Student Self Evaluation Form
Sixth Grade
Page 3

BEST COPY AVAILABLE

Student Self-Evaluation Form, Sixth Grade, p. 4

I would like to learn to clean my cuby

To be quicker on easy math problems

learn more soccer skills

learn how to make an outline

Help me calm down sometimes at music

Shutesbury Elementary School
Student Self Evaluation Form
Sixth Grade
Page 4

WAIALAE ELEMENTARY SCHOOL
Grade 1 Learning Profile
1994-95 School Year

Student:______________________ Room____ Teacher:______________

E = Emergent B = Beginning D = Developing Ex = Expanding M = Mature	1 = 1st Quarter 2 = 2nd Quarter 3 = 3rd Quarter 4 = 4th Quarter

ATTENDANCE	1	2	3	4
Days Present				
Days Absent				
Times Tardy				

READING	Emergent	Beginning	Developing	Expanding	Mature
Shows an interest in reading	E	B	D	Ex	M
Uses picture cues	E	B	D	Ex	M
Uses phonetic cues	E	B	D	Ex	M
Self corrects	E	B	D	Ex	M
Understands what is read	E	B	D	Ex	M
Selects and reads books with confidence	E	B	D	Ex	M
Reads in a variety of areas	E	B	D	Ex	M

WRITING	Emergent	Beginning	Developing	Expanding	Mature
Shows an interest in writing	E	B	D	Ex	M
Able to self select topic	E	B	D	Ex	M
Able to express ideas clearly	E	B	D	Ex	M
Able to expand and add details	E	B	D	Ex	M
Writes for a variety of purposes	E	B	D	Ex	M
Shares writing with others	E	B	D	Ex	M
Uses conventions of writing	E	B	D	Ex	M
Writes legibly	E	B	D	Ex	M

MATHEMATICS			
Demonstrates number sense	Not yet	Developing	Strong
Displays an understanding of math concepts	Not yet	Developing	Strong
Uses a variety of problem solving strategies	Not yet	Developing	Strong
Communicates mathematically	Not yet	Developing	Strong
Willing to explore and experiment	Not yet	Developing	Strong

Waialae Elementary School
Grade 1 Learning Profile
Page 1

BEST COPY AVAILABLE

Handout A4.5, H5, p. 1

LISTENING/SPEAKING

Listens attentively	Not yet	Developing	Strong
Expresses ideas	Not yet	Developing	Strong
Discusses/comments appropriately	Not yet	Developing	Strong
Willing to share ideas	Not yet	Developing	Strong

UNIT STUDY (Social Studies/Science)

Demonstrates an understanding of concepts	Not yet	Developing	Strong
Participates in discussions and activities	Not yet	Developing	Strong
Shares materials related to unit of study	Not yet	Developing	Strong
Shows a willingness to extend learning	Not yet	Developing	Strong

PERSONAL DEVELOPMENT

Accepts learning challenges (risk taking)	Not yet	Developing	Strong
Is developing self-confidence	Not yet	Developing	Strong
Participates in class activities	Not yet	Developing	Strong
Shows a willingness to explore and discover	Not yet	Developing	Strong
Follows directions	Not yet	Developing	Strong
Demonstrates responsibility	Not yet	Developing	Strong
Shows consideration for others & their ideas	Not yet	Developing	Strong
Works cooperatively in groups	Not yet	Developing	Strong
Works independently	Not yet	Developing	Strong
Demonstrates perseverance	Not yet	Developing	Strong
Makes appropriate use of time and materials	Not yet	Developing	Strong
Attempts to solve problems in a variety of ways	Not yet	Developing	Strong
Can negotiate conflicts	Not yet	Developing	Strong
Respects rules and regulations	Not yet	Developing	Strong
Cares for personal belongings	Not yet	Developing	Strong
Cares for property belonging to others	Not yet	Developing	Strong
Accepts suggestions/corrections	Not yet	Developing	Strong

PHYSICAL EDUCATION

Participates in physical education activities	Not yet	Developing	Strong
Demonstrates sportsmanship	Not yet	Developing	Strong

MANDARIN

Is building an understanding of the language and culture	Not yet	Developing	Strong

HAWAIIANA

Is building an understanding of the language and culture	Not yet	Developing	Strong

Waialae Elementary School
Grade 1 Learning Profile
Page 2

BEST COPY AVAILABLE

Waialae Elementary School Writing Continuum and Parent Letter

DRAFT

Waialae Elementary Writing Continuum K-2

Emergent	Beginning	Developing	Expanding	Mature
• Reproduces letter-like forms or actual letters. • Recognizes some letters of the alphabet • Is able to use some knowledge of letter/sound associations for invented spelling of words. • Attempts to reread what has been written.	• Able to recognize all letters of the alphabet • Able to use upper and lower case letters appropriately • Able to analyze and record consonant sounds and easy to hear vowels • Places spaces between words • Able to reread what has been written	• Uses random invented spelling • Writes with increasing fluency • Writes from experience • Rereads for meaning • Able to add detail • Begins to edit • Enjoy sharing and responding to others' writing	• Showing more independence in writing • Revises and extends story • Able to do some proofreading and editing skills • Begins to use various punctuation and capitalization appropriately	• Includes more story elements - choice, setting, problem, solution • Provides more information to support main idea • Uses consistent developmental invented spelling • Uses paragraphs for organization • Uses editing and revising skills • Writes on a variety of topics

Waialae 1/95v.1

February 1995

Dear Parents,

On the Learning Profile you will notice a continuum that shows your child's development in reading and writing. It is labeled with the descriptors: Emergent, Beginning, Developing, Expanding and Mature. To help you understand the terminology, we have created a chart which lists behaviors of each characteristic level. It is our intent to give you a clearer picture of your child's progress. These charts are in draft form and are still works in progress.

We hope you will find this reporting instrument useful in understanding your child's progress in school.

Sincerely,
First Grade Teachers

BEST COPY AVAILABLE

Waialae Elementary School
Reading Continuum

DRAFT

Waialae Elementary Reading Continuum K-2

Emergent	Beginning	Developing	Expanding	Mature
• Enjoys being read to • Looks at books alone and with other • Is curious about print • Begins to experience "book sense" (front to back, right side up) • Joins in the oral reading of familiar stories, especially those with rhythmic language and patterns • Pretends to read • Uses pictures to help tell the story • Begins to predict what may happen next • Knows some letters and is gaining awareness that letters have sounds • Recognizes some words	• Chooses to read by him/herself • Reads a simple book in which language patterns are repeated • Reads known and predictable books with growing confidence, but still needs support with new and unfamiliar ones • Remembers sequence of events • Understands that the printed words, as well as the pictures convey meaning • Uses pictures to gain meaning from the material • Relates personal experiences to text with adult support • Begins to self correct using pictures and context • Identifies and names most letters and sounds • Recognizes common words from the world around them • Beginning to develop sight vocabulary from text	• Chooses books independently, reads alone and with others • Shows increasing knowledge of print conventions • Takes risks and makes guesses • Uses pictures to predict words • Makes predictions using pictures, phonics and context • Uses language cues to understand words • Begins to self-correct when reading • Rereads to regain meaning • Relates personal experiences to text and discusses ideas • Knows how to uses reading materials to get information • Uses some letter-sound relationships • Reads with enough fluency to focus on meaning of what is read • Is developing some oral fluency • Increasing basic sight vocabulary	• Selects and reads books with confidence • Chooses from a wide variety of written materials • Shows understanding of print conventions • Reads books that have more complex language and story lines • Understands that taking risks and making guesses are essential parts of reading • Beginning to use comprehension strategies independently (sequence events, backs up statements with proof from reading, connects own experience with books) • Self-corrects when miscues disrupt meaning' • Reads for a variety of purposes • Gathers information and ideas from a wide range of texts • Uses ideas from reading as tools for solving problems or for further learning • Summarizes key ideas • Reads with oral fluency • Reads with expression	• Reads a wide range of materials • Reads to learn rather than learning to read • Comprehends at different levels (literal, strict meaning, interpretive, inferential, critical) • Makes thoughtful and discriminating personal responses to a range of literary texts • Uses strategies of predicting, confirming and self-correcting, quickly, confidently and independently • Builds flexible reading speeds which-vary depending on purpose • Reads with increasing oral fluency • Reads with increasing expression

Walalae 1/95v.1

BEST COPY AVAILABLE

CAROLYN SCHOELLER - THIRD GRADE

At Crow Island School, students in second through fifth grades experience a "Portfolio Night" with their parents at school. Second and third grade children and parents participate in this experience in February, and fourth and fifth grade children and parents attend in May. Younger children have already begun compiling their own portfolios in kindergarten and first grades and will look forward to Portfolio Nights in future years.

Just What Is Portfolio Night?

- a special evening for Mom, Dad, and child together
- an intimate, private family time (no siblings attend)
- a very important, exciting evening to the child
- an informal, comfortable atmosphere
- child explores work samples with his/her parents
- child showing and teaching parents what he/she has learned
- child assumes the teacher role, while parents assume the learner role
- child discovers and recognizes personal growth. (This recognition of progress often surprises the child him/herself.)
- a time of sharing
- a sense of great accomplishment (Child looks at work done earlier in the year, as though he/she had done it two years ago!)
- child is "in charge"
- child is eager to display maturity
- child directs the evening, does the explaining
- personalized, individualized
- self-assessment by the child
- a time for communication
- child and parents interacting (teacher available for assistance)
- a time for reflection on learning and progress
- a time for setting goals together
- a child's feeling of ownership and pride in his/her work

Portfolio Night is a very unique way of reporting progress to both children and parents!

Crow Island School
What is Portfolio Night?

BEST COPY AVAILABLE

638

Handout A4.5, H6, p. 1

PORTFOLIO NIGHT - 3RD GRADE

February, 1994
(Posted on Board for the Children)

1. Share your Portfolios with your parents.
2. See how much work you've done this year.
3. Share what you've learned.
4. Teach your parents some new things.
5. Explore our room together.
6. See the videos. Remember the Xap Shots too!
7. Write your Personal Goals with your parents.

SOME PORTFOLIO NIGHT IDEAS AND ACTIVITIES

February, 1994

(Each child had a copy of this at his/her "station". They could pursue any or all of these activities after looking through their Portfolios)

Sunrise/Sunset
Diameter/ Circumference
Español
Colonies
Poetry/ Biographies

Highs/Lows
Square Feet/ Square Yards
Acrostics
Whirly Birds
Counters

Geometric Prisms (faces, edges, vertices)
Tape Measures/ Measuring Sticks (US Customary/Metric)
Class Writing Titles (Halloween/ Thanksgiving, Holiday
Circuit Puzzles (Testers)

Perimeter/Area
Calculators
Miss Pickett Stories
Math Games (Ups, Downs, Castoff, Math Down)

Packages (Length/Girth, Special Handling Requirements)
Slates
Research Books
Special Focus Books

Crow Island School
Portfolio Night
Ideas and Activities

BEST COPY AVAILABLE

PORTFOLIO NIGHT: GUIDELINES - 5TH GRADE

THESE ARE GUIDELINES TO HELP YOU AS YOU CONFERENCE WITH YOUR PARENTS ON PORTFOLIO NIGHT.

Goals: You have set many goals for yourself this year. Discuss how well you have achieved YOUR goals. In what areas have improved: In what areas do you feel you need to continue to "work" to improve?

What are your strengths? Weaknesses? What can you do this summer to work and improve in these areas?

Discuss with your parents what your goals might be for 6th grade this fall.

Math:. In what areas of math do you think you have done well? In what areas have you struggled?

What can you do this summer to keep your skills going? How can your parents help with this? What can you do to help yourself?

HINT: Don't ignore the books you've completed this year. Perhaps review some of them over the summer.

Spelling What is the process we use in class each week? Explain why our class takes a "pretest" each week.
English

How well do you proofread now for spelling and grammar? For ;you, personally, has there been an improvement? Ask your parents what they think? Look over your spelling, reading and creative writing as you discuss this. Do you feel that this needs to be a BIG focus for next year?

Have ;you improved this year?

Reading: Our class has read Racing the Sun, The Cay, Where the Red Fern Grows, Cry of the Crow, the Big Wave, A Wrinkle in Time, Escape from Warsaw:. Every three weeks you have read a book in a different "genre" and created a book project about the book. Which book project did you enjoy the most. Favorite author? Why?

Do you feel your reading has improved? Do you enjoy personal reading?

What kind of reading do you plan to do this summer?

How have your comprehension questions improved from the very first book? Look over the work carefully with your parents.

Compare 4th grade reading packets with 5th grade reading papers. What difference/improvements do you notice?

Science: You have worked on Energy Sources, (Invention Convention) and The Human Body (skeletal, muscles & the reproductive system).

what did you like about each of these units? What challenged your thinking? Which units helped you to understand how patient a scientist must be before he/she comes up with results?

Social Studies: Our class has learned a lot about world issues and concerns through "World Newsmap" and through class discussions. We have looked at our environment through the rain forest unit and reading Jean Craighead George's book on a day in a rain forest.

Tell your parents about the "Endangered Species" quilt the entire 5th grade is making and why. Explain where the quilt will be on display. Tell why you feel it is important to "make a statement" about our environment at your age.

Diversity has been a major theme at Crow Island this year. Our class has learned about Africa and Japan.

Talk about the importance of diversity and understanding of other cultures.

Research Project... This end of the year project began with mini-projects, followed by CREATE-A-KIT (explain and show photographs), and now the research project. You are using a "process" approach.

Crow Island School
Portfolio Night:
Guidelines - 5th Grade
Page 1

BEST COPY AVAILABLE

Discuss the steps that you have used and are using: reading, visiting the public library for research, note-taking, writing rough drafts from notes, proofreading first by yourself and then classmates and parent volunteers, keeping your bibliography cards up to date, etc.

Discuss how this process works for you. Were you able to meet all the deadlines so far? Do you place the deadline calendars in places you would be sure to look?

What are specific problems writing the research paper that you need to work on next year?

If everything has been right on schedule for you, what have you personally done to make it this way? Discuss with your parents.

Creative Writing: We have written daily. Compare samples of your writing from 4th grade, 3rd grade, etc. (use your archive) with your 5th grade writing.

Look at the content, grammar, sentence and paragraph structure and appearance. How has your writing changed? For the better? Worse? Why?

What are your favorite "pieces" you have written this year? Read sections of them to your parents. Young Author's Conference?

The more you write, the better you write. Just think, when you write in your journals this summer, no one will "proofread" it!!!

Art: Share your huge "portfolio" with your parents. What is your favorite art project? Why? Which of your many intelligences are at work as you create these projects?

Music: Band/orchestra/general music/chorus/J.S. Bach? How does music make you feel? What have you learned this year? What does music help to teach you about structure and discipline? What do you enjoy about music? What composers have you learned a lot about this year? Your favorite? If our crew video-taped you performing, did you listen critically to your playing? What passages you played could be improved? How did most of those difficult sections you practiced over and over sound to you? Were you pleased with your overall performance?

Physical Education: What have you learned this year that you feel has improved your physical fitness? What has been the most difficult area for you? In which activity are you the most successful? Do you cooperate in team sports, or is that an area that needs improvement?

Computers/Technology: Do you think you have improved your keyboarding skills this year? Why? Do you think you write more because you have good keyboarding skills?

Look over your various rough drafts of your research project. What have you learned? Do you feel that you write more because you are using the computer.

Video-taping (if this applies to you): what have you learned from this responsibility? Do you want to continue this at Washburne?

Anything else? Remember to spend some time looking through your "archive" with your parents.

Crow Island School
Portfolio Night:
Guidelines - 5th Grade
Page 2

BEST COPY AVAILABLE

NAME______________
DATE______________

REFLECTIONS ON YOUR PORTFOLIO EVENING: May 27, 1993

1. How did the Portfolio Night Guidelines help to keep your parents focused on what you were discussing with them?

____
____
____

2. What sections of your Portofolios did you spend the most time looking at and discussing with your parents?

____
Why?____
____
____

3. Did your "archive"(if you have one) help you when reflecting on a. what you have learned? b. how much you have improved in some areas c. areas that need work on an on going basis.

____
____
____
____
____

4. What surprised your parent/parents about your portfolio?____________
____
____
____
____

5.Think of an example of a piece of work that you found in your "archive" and try to remember how you worked on it then--do you still work the same way on a project? How have you grown in certain areas?

____
____
____

6. Looking back over your work, what do you consider the areas that you have improved considerably? Why? What areas have you put in the most work? What are your weaknesses? Are they the same as when you started at

Crow Island School
Relections on Your
Portfolio Evening
Page 1

BEST COPY AVAILABLE

C.
I.?____

7. When you watched the Portfolio Evening video -what did you think about? What were some of your favorite projects throughout the year? Why? Did the video help you to remember things you had forgotten that you had worked on?

8.Since you were"in charge" of last night with your parent/s-what do you feel they were impressed with that you had accomplished in 5th grade?____

9. How comfortable did you feel organizing ang conferencing with your parents? Would you have done anything differently?

10 What did you enjoy about last night?Would you change anything?____

11. Your personal suggestions for Portfoloio Evening (next year) would be greatly appreciated.
(suggestions)

Crow Island School
Relections on Your
Portfolio Evening
Page 2

BEST COPY AVAILABLE

SCHOOL DISTRICT #62 (SOOKE)

WISHART ELEMENTARY SCHOOL

Student Progress Report

Student: [redacted]
Div: 12 Gr: 1
Teacher: Mrs. Davidson, Mrs. Politano
6 yr 9 mo

Reporting Period June 95
(month) (year)

[redacted] loves to read and is keen to try challenging books. She is determined to figure out what is written and uses a variety of skills such as sounding out, picture cues, recognizing words she knows and making sense from the rest of the text. She is now able to read books such as "Clifford and the Grouchy Neighbors" and simply needs practice and opportunities to build upon and refine her reading abilities.

She writes with confidence and is now able to write messages that others can read. She still tends to include some letters that don't fit the sound symbol match but with practice and participation in activities such as C.O.P.S. and "Mystery Message" she will become more accurate. She impressed us all when she noted the "u" as the missing letter in "our" in a close proofreading activity. The next step to refine [redacted] writing would be to have her work on letter formation and organization of the message on the page.

[redacted] artistic talent has shone throughout the year. She puts a lot of effort into making interesting, colourful and imaginative things. When she is busy on a project [redacted] is very focussed and uses her time well.

[redacted] clearly understands and can demonstrate what she knows about primary math concepts. She can add and subtract numbers easily and will be doing more regrouping with larger numbers in the fall. She takes the time to print her numbers clearly and is organized with materials she uses.

Consistent positive behaviour continues to be a major challenge for [redacted] to maintain. She has shown growth in her ability to share attention in a group but has had many incidents of hitting, biting and throwing rocks at other children. She also makes remarks that put her in conflict with others. Although every situation has two sides [redacted] does seem to initiate or react violently in negative situations. We have tried to help her see the consequences of her behaviour and appreciate your willingness to work together to help her make more effective choices. At this stage we are concerned that she is alienating herself from others and that she is entering her third year at school very much in need of developing ways to get along with others. Our goal is to begin next year with a well defined plan and support to help [redacted] develop the skills she needs to get along in a classroom and on the play ground and to make a productive contribution to her learning community.

Wishart Elementary School
Student Progress Report, Grade 1

BEST COPY AVAILABLE

SCHOOL DISTRICT #62 (SOOKE)

WISHART ELEMENTARY SCHOOL

Student Progress Report

Student: [redacted]
Div: 12 Gr: 2
Teacher: Mrs. Davidson, Mrs. Politano
7 yr 9 mo

Reporting Period: June '95
(month) (year)

This report will contain a description of:

a) what the student is able to do,
b) the areas requiring further attention or development, and
c) ways of supporting the student's learning.

Included are comments that describe student behavior (attitudes, work habits and effort).

June 1995

[redacted] has had a most successful year. He has matured and has become a very dependable and helpful person.

He can read with confidence. He is always willing to use his skills to help others. When asked to pick a poem he hadn't read before [redacted] chose "Who I Am" by Felicia Holman. He could read the whole thing without a problem and could tell about what he had read. He is enjoying the chapter book "Horrors of the Haunted Museum. At this stage all he needs are opportunities to read so he can polish his skills.

His ability to express himself through writing is excellent. The day he took two full pages of numbered, specific notes about the good things happening in our class was a perfect example of his competence. He enjoys writing and can be counted on to put a lot of effort into doing a good job. His spelling is becoming more standard and he likes to try to spell words and use conventions such as "tion" "ing" and apostrophes. He tends to print everything in capital letters so the next logical step is to move to appropriate use of capital and lower case letters.

He approaches the arts and physical acitivities with the same excellent attitude he brings to his other work. His willingess to be thoughtful and put in the necessary effort have brought him success in all areas of his school life.

[redacted] interest and enthusiasm for math and sciences shows his ability to use numbers and understand concepts quickly and accurately.

Because [redacted] has been successful in all areas of the Primary Program, meeting his needs is a matter of continuing to help him build upon his skills and follow his interests.

Having [redacted] in our class has made it a nicer place to be. We wish him continued success in school.

Assigned to Grade 3.

C Politano

B. J. Esch

Student Progress Report, Grade 2

645

Handout A4.5, H7, p. 2

BEST COPY AVAILABLE

Travis Heights Elementary School Progress Reports

Travis Heights Elementary is a K-5 multi-age school (students of 2 or 3 different grade levels attend classes together). The school uses a 9-weeks grading cycle. Parental reporting of progress occurs four times annually and takes one of two forms:

- Parent conferences the first and third quarters and
- Narrative reports the second and final quarters.

The only letter or percentage scores reported to the parent, school, district, or state are those of national and standardized tests. These scores are not included in student progress reports.

The first conference acquaints parents with school subject area and developmental benchmarks students are expected to attain by the fourth grade. An optional parent evaluation of the session is encouraged and a form is provided for this purpose.

The second quarter, narrative reports are sent to the parents. In addition to a form for core subject areas, each special area sends out a separate form (see attached). The form includes a section for student self-assessment as well as a section for parents to assess their child's performance.

The third conference is a three-way conference (student/teacher/parent) led by the student and focused on individual portfolios.

The final narrative report is sent home two days before the end of the semester to allow time for parental assessment comments.

Travis Heights Elementary
Progress Report Procedure

Handout A4.5, H8, p. 1

Austin Independent School District
Travis Heights Elementary
Principal, Marilyn Butcher
19__/ 19__
Narrative Progress Report

For reporting period
1 2 3 4

Special Programs
___ ESL ___ Resource
___ Bilingual ___ Speech
___ Content Mastery

Student____
Teacher____
Instructional Team ______________________(Grade ________)

Integrated Language Arts / Social Studies

Mathematics / Science / Health

Personal Development

Reporting Period 2: White Copy – parent Pink Copy – teacher
Yellow Copy – parents sign & return to school – Thank you!

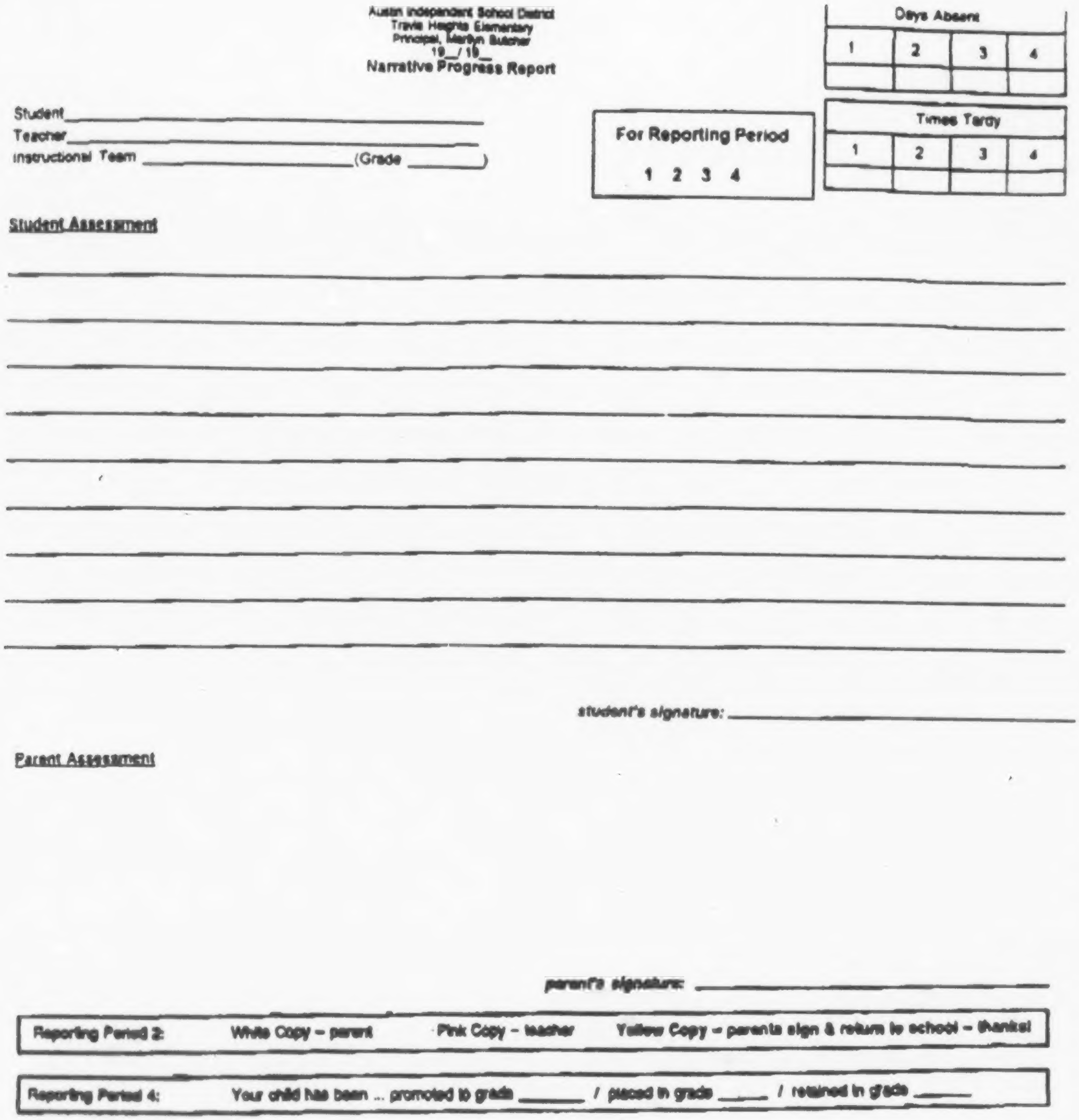
parent signature:________________________________

Travis Heights Elementary
Narrative Progress Report
Page 1

647

Handout A4.5, H8, p. 2

BEST COPY AVAILABLE



Austin Independent School District
Travis Heights Elementary
Principal, Marilyn Butcher
19__/19__
Narrative Progress Report

Days Absent			
1	2	3	4

Student ______________________________

Teacher ______________________________

Instructional Team ________________(Grade ________)

For Reporting Period

1 2 3 4

Times Tardy			
1	2	3	4

Student Assessment

student's signature: ______________________________

Parent Assessment

parent's signature: ______________________________

Reporting Period 2: White Copy – parent Pink Copy – teacher Yellow Copy – parents sign & return to school – thanks!

Reporting Period 4: Your child has been ... promoted to grade ______ / placed in grade _____ / retained in grade ______

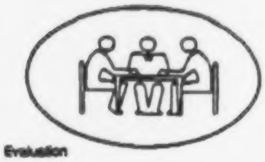
Travis Heights Elementary
Narrative Progress Report
Page 2

Handout A4.5, H8, p. 3

BEST COPY AVAILABLE

Travis Heights Elementary
Conference Report

Parent 3-way



Student:______________________________

Teacher:______________________________ ☐ S. S. P.

Team:______________ Date:____________ ☐ District Conference Evaluation

Grade: ______

Summary of Conference Discussion

Strengths

Focus Areas

Action Plan

Goal(s):

Action to be taken by

student:

parent:

teacher:

____
student's signature
(3-way conference only)

____
teacher's signature

____
parent's signature

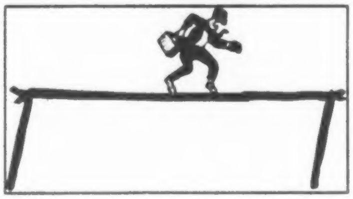
Travis Heights Elementary 3-Way Parent Conference Report

Handout A4.5, H8, p. 4

BEST COPY AVAILABLE

Activity 4.6

How to Convert Rubric Scores to Grades



Purposes:

1. To explore the issues involved when converting rubric scores to grades
2. To apply knowledge about grading and the role of rubrics in standards-based education to a specific case

Uses:

This activity is for an audience advanced in their study of classroom assessment. Prerequisites include: (a) knowledge about the reasons for changes in assessment and the place of assessment in standards-based instruction (see **Chapter 1** text and activities); (b) a broad view of grading and reporting concerns and issues (**Activity 4.2**—*Putting Grading and Reporting Questions in Perspective* or **Activity 4.3**—*Grading Jigsaw*); and (c) **Activity 4.8**—*Guidelines for Grading* (which establishes that the main purpose for grading is communication and provides an opportunity to discuss sound grading practices).

Rationale:

Teachers want more than general discussions of issues surrounding grading and reporting. They want solutions to the situations that press them on a daily basis. One of these situations is the need to reconcile innovative student assessment designed to be standards-based and influence instruction (such as the use of rubrics to teach skills and track student performance) with the need to grade. This activity gives teachers a chance to actually make some decisions and plan a course of action.

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens
- **Overheads: A4.6,O1**—*How to Convert Rubric Scores to Grades Purposes,* **A4.6,O2**—*Six-Trait Model,* and **A4.6,O3**—*Discussion Questions*
- **Handouts: A4.6,H1**—*Six-Trait Model,* **A4.6,H2**—*Instructions,* **A4.6,H3**—*Memo*

Time Required:

60-75 minutes

Facilitator's Notes:

1. (2 minutes) Introduce the purposes of the activity with **Overhead A4.6,O1**—*How to Convert Rubric Scores to Grades Purposes.*

2. (15-20 minutes) **Context setting.**

   Participants will read a memo written by a district assessment coordinator in response to a request for help from teachers who were attempting to reconcile their use of a writing rubric for instruction with their need to grade. The specific situation was that teachers had been using the six-trait model to teach writing and keep track of student progress. At the end of the grading period they said, "OK, we did six-trait. Now, how do we grade?"

   (Note: Although we use the Six-Trait Model in this case study, the same issues will arise with any multi-trait rubric.)

   In order to understand the memo written in response to this question, participants need to know a few things:

   - The district instituted use of performance assessment in writing for its positive influence on instruction and its ability to track student skill levels in a more useful fashion. The philosophy is this: If we clearly define what quality writing looks like and illustrate what we mean with samples of student writing, everyone (teachers, students, parents, etc.) will have a clearer view of the target to be hit. And, indeed, the process of determining characteristics of quality and finding samples of student work to illustrate them (and teaching students to do the same) improves instruction and helps students achieve at a higher level.

   - The characteristics of quality writing used by the district is embodied in a writing rubric called the "Six-Trait Model." Since the memo refers to the six traits, participants need to know a little about them. **Overhead A4.6,O2** and **Handout A4.6,H1** describe the six-trait model. (The entire six-trait rubric is included as **Sample A.10**—*Six-Trait Model* in **Appendix A**—*Sampler*.) The six-trait model should cover the same features of quality as other writing assessment methods used by participants. What tends to be different about the six-traits is the clarity with which they are written, the availability of student-language versions, and the emphasis on use in instruction.

     Each of the six traits is scored on a scale of one to five, where five is high. These are *not* intended to correspond to "A, B, C, D, and F." Teachers are encouraged not to use the traits in a lock-step fashion. Rather, they should use the traits that make sense in any given instructional instance, to weight them differently depending on the situation, and ask students to add language that makes sense to them.

- Provide a brief overview of the methods in the memo for converting rubric scores to grades. The first and fourth methods are essentially logic processes. They don't require adding numbers at all. Method one essentially says that if the student has various patterns of scores, they should get thus and so grade. Several possible patterns and their relationship to grades are proposed. Method four asks that the teacher come up with rules for converting rubric scores to "percents." For example, an "A" might be equivalent to a score of "4" so "4's" would get 90%.

  The middle two methods require adding, multiplying and dividing numbers. For example, the second method requires adding up all the student's earned points and comparing it to the total possible points (e.g., 90%=A). A variation on this (method three) weights some traits more than others before comparing earned points to total possible points.

3. (5 minutes) **Set up the activity**.

Participants will read and discuss the attached memo, **Handout A4.6,H3**—*Memo*, outlining several methods for converting rubric scores to grades. Then they will answer the discussion questions on **Overhead A4.6,O3** and **Handout A4.6,H2**—*Instructions*:

a. (Optional) According to Kohn, (see **Activity 4.2**) what level of question is: "We want to strengthen grading by improving the assessment on which grades are based; how do we combine the numbers at the end to come up with a grade?

   (The original question posed by the teachers corresponds to Level I/II as proposed by Kohn. They wanted to assess for purposes other than grading [continuous monitoring of student progress and student self-assessment], but then they also needed to combine the numbers at the end to come up with a grade. The goal in this activity is to encourage participants to solve this Level I/II question with Level III arguments. At Level III teachers question the entire enterprise of assessment and grading—why do we assess and why do we grade, and do our current procedures accomplish our purposes?)

b. In this district, the purpose for using the rubric scoring scheme was motivating students, providing feedback (accurate communication of achievement in relationship to standards), encouraging students to self-assess, and using assessment to improve achievement. Although the primary purpose for grading should be accurate communication of student learning, might there be hidden purposes for grading that might conflict with the instructional purposes for the rubric?

c. What are the advantages and disadvantages of each method for converting rubric scores to grades if the purpose for grading is to accurately report student achievement status and support the other classroom uses of the rubric?

d. Which method would you recommend and why?

e. How does this compare with current practice at your school? How might this change after today's discussion?

Note that there are no single correct answers to these questions. The goal is to have a sound rationale for the procedure chosen, regardless of what it is.

4. (30-45 minutes) Ask participants to read the memo and discuss the options. The following represent the discussions that teachers engage in when attempting to choose a defensible method. (Many of these discussion points raise Level III concerns.)

- If the main purpose of grading is accurately communicating student achievement, the best method would be 1 or 4 because dry conversion of scores to percents masks important information, like whether a student has improved over time. Why not look at only the most recent work?

- If we choose different methods, the same student could get different grades. Should we have a common procedure across teachers?

- Every assessment should be an opportunity to learn. Which method would best encourage this goal? Probably, the one where students choose the work to be graded.

- We have to be really clear on what the grade is about—how will it be used and what meaning will it have to students and parents. So different methods might be used at elementary and secondary.

- Use of rubrics in the classroom and grading come from different philosophies about the role of assessment in instruction and learning. Many times grades are just used to sort students without having a clear idea of what the grade represents—achievement, effort, helping the teacher after school or what. If we have to grade, we need to select the method that keeps best to the spirit of the rubric.

- Methods 2 and 3 can be misleading because a "3" corresponds to 60% which is often seen as an "F." Yet the description for "3" doesn't seem to indicate failing work. Are we standards-based or not?

- There is general dislike of the total points and weighting methods because they don't keep as much to the spirit of rubric use in the classroom or having the grading being an episode of learning. Groups tend to like methods that give them flexibility and encourage learning—the option to place more weight on papers produced later in the grading period or requiring students to choose the papers on which to be graded and providing a rationale for their choice.

5. (10 minutes) Get back together and discuss **Handout A4.6,H2**, items 4 and 5.

How to Convert Rubric Scores To Grades

Purposes:

1. To explore the issues involved when converting rubric scores to grades

2. To apply knowledge about grading and standards-based education to a specific case

Overhead A4.6,O1

Six-Trait Model

Ideas—the heart of the message

Organization—the internal structure of the piece

Voice—the tone and flavor of the author's message

Word Choice—the vocabulary a writer chooses to convey meaning

Sentence Fluency—the rhythm and flow of the language

Conventions—mechanical correctness

Overhead A4.6,O2

Discussion Questions

1. (Optional) According to Kohn, what level of question is: "We want to strengthen grading by improving the assessments on which grades are based; how do we combine the numbers at the end to come up with a grade?
2. Although the primary purpose for grading should be accurate communication of student learning, might there be hidden purposes for grading that might conflict with the instructional purposes for the rubric?
3. What are the advantages and disadvantages of each method in the memo if the goal of grading is to accurately communicate achievement and support the other classroom uses of the rubric?
4. What method would you select? Why?
5. How does this compare with current practice at your school? How might this change after today's discussion?

Overhead A4.6,O3

656

The Six-Trait Model
In Brief[1]

Ideas—Ideas are the heart of the message, the content of the piece, the main theme, together with all the details that enrich and develop that theme. The ideas are strong when the message is clear, not garbled. The writer chooses details that are interesting, important, and informative—often the kinds of details the reader would not normally anticipate or predict. Successful writers do not tell readers things they already know (e.g., "It was a sunny day, and the sky was blue, the clouds were fluffy white..."). They notice what others overlook, seek out the extraordinary, the unusual, the bits and pieces of life that others might not see.

Organization—Organization is the internal structure of a piece of writing, the thread of central meaning. It doesn't matter what the pattern is, so long as it fits the central idea well. Organizational structure can be based on comparison-contract, deductive logic, point-by-point analysis, development of a central theme, chronological history of an event, or any of a dozen other identifiable patterns. When the organization is strong, the piece begins meaningfully and creates in the writer a sense of anticipation that is ultimately, systematically fulfilled. Events proceed logically; information is given to the reader in the right doses at the right times so that the reader never loses interest and never the "big picture"—the overriding sense of what the writer is driving at. Connections are strong, which is another way of saying that bridges form one idea to the next hold up. The piece closes with a sense of resolution, tying up loose ends, bringing things to closure, answering important questions while still leaving the reader something to think about.

Voice—The voice is the writer coming through the words, the sense that a real person is speaking to us and cares about the message. It is the heart and soul of the writing, the magic, the wit, the feeling, the life and breath. When the writer is engaged personally with the topic, he/she imparts a personal tone and flavor to the piece that is unmistakable his/hers alone. And it is that individual something—different from the mark of all other writers—that we call voice.

Word Choice—Word choice is the use of rich, colorful, precise language that communicates not just in a functional way, but in a way that moves and enlightens the reader. In good descriptive writing, strong word choice paints clear pictures in the reader's mind. In

Handout A4.6,H1, p. 1

---

[1] From *The Six Traits*, 1997, Vicki Spandel and Ruth Culham, Northwest Regional Educational Laboratory.

good expository writing, strong word choice clarifies and expands ideas. In persuasive writing, careful word choice moves the reader to a new vision of things. Strong word choice is characterized not so much by an exceptional vocabulary that impresses the reader, but more by the skill to use everyday words well.

Sentence Fluency—Sentence fluency is the rhythm and flow of the language, the sound of word patterns, the way in which the writing plays to the ear—not just to the eye. How does it sound when read aloud? That's the test. Fluent writing has cadence, power, rhythm and movement. It is free of awkward word patterns that slow the reader's progress. Sentences vary in length and style, and are so well crafted that reading aloud is a pleasure.

Conventions— Conventions are the mechanical correctness of the piece—spelling, grammar and usage, paragraphing (indenting in the appropriate spots), use of capitals, and punctuation. Writing that is strong in conventions has usually been proofread and edited with care. Handwriting and neatness are not part of this trait. The key is this: How much work would a copy editor need to do to prepare the piece for publication?

Handout A4.6,H1, p. 2

Instructions: "How to Convert Rubric Scores to Grades"

There are no right or wrong answers to these questions. What is important is to have a good, defensible reason for whatever is chosen.

1. (Optional) According to Kohn, what level of question is: "We want to strengthen grading by improving the assessments on which grades are based; how do we combine the numbers at the end to come up with a grade?"

2. In this district, the purpose for using the rubric scoring scheme was motivating students, providing feedback (accurate communication of achievement in relationship to standards), encouraging students to self-assess, and using assessment to improve achievement. Although the primary purpose for grading should be accurate communication of student learning, might there be hidden purposes for grading that might conflict with the instructional purposes for the rubric?

3. In groups, read and discuss the memo and written material on converting rubric scores to grades. What are the advantages and disadvantages of each option described?

Option	**Advantages**	**Disadvantages**
1		
2		
3		
4		

4. What advice would you give a teacher? Why?

5. How does this compare with current practice at your school? How might this change after today's discussion?

Handout A4.6,H2

February 7, 1997

To: Teachers

From: Linda L. Elman,
Testing Coordinator
Central Kitsap School District, Silverdale, Washington

RE: Converting Rubric Scores for End-of-Quarter Letter Grades

Introduction

There is no simple or single way to manipulate rubric scores so that they can be incorporated into end of quarter letter grades. This paper contains a set of possible approaches. Or, you may have developed a process of your own. Whatever approach you choose to use, it is important that you inform your students about your system. How grades are calculated should be open to students rather than a mystery. In addition, you need to make sure that the process that you use is reasonable and defensible in terms of what you expect students to know and be able to do as a result of being in your class.

In all cases, you might not want to use all papers/tasks students have completed as the basis for your end of quarter grades—you might choose certain pieces of student work, choose to emphasize certain traits for certain pieces, let students choose their seven "best" pieces, etc. You might only want to score certain traits on certain tasks.

You might consider placing most emphasis on works completed late in the grading period. This ensures that students who are demonstrating strong achievement at the end of a term are not penalized for their early "failure." It also encourages students to take risks in the learning process. Whatever you choose to do, you need to have a clear idea in your mind how it helps you communicate how students are performing in your classroom. In the end, what you need to have are adequate samples of student work that will allow you to be confident about how well students have mastered the skills that have been taught. (Do you have enough evidence to predict, with confidence, a student's level of mastery on their next piece of work?)

Down the road we will want to convene a group of teachers to come up with a common acceptable and defensible system for converting rubric scores to grades. In the short run here are several methods which can be used to convert rubric scores to letter grades.

Handout A4.6,H3, p. 1

Methods:

The methods described here can be used with any tasks, papers, or projects which are scored using rubrics. The example used is from writing assessment, but the methods identified here are not restricted to writing.

In **Table 1** we have Johnny's scores on the 5 pieces of writing we agreed to evaluate this term on all six traits.

Table 1: Johnny's Writing Scores on Five Papers

Johnny's Scores	Ideas and Content	Organization	Word Choice	Sentence Fluency	Voice	Conventions	Total
Paper 1	3	2	2	3	1	4	15
Paper 2	4	2	3	4	3	4	20
Paper 3	5	4	5	5	2	3	23
Paper 4	4	4	4	4	2	4	23
Paper 5	5	5	5	5	4	4	28
Totals	21	17	19	21	12	19	109

METHOD 1: Frequency of Scores Method. Develop a logic rule for assigning grades. The following are just four of many possible ways you could go about setting up a rule for assigning grades in writing.

- To get an A in writing you have to have 50% of your scores at a 5 with no scores of Ideas and Content, Conventions, or Organization below 4.
- To get a B you have to have 50% of your scores a 4 or higher with no Ideas and Content, Conventions or Organization below 3, and any other score below a 3 counterbalanced by a score of 4 or higher. Etc.

O R

In this class, in writing,

- Mostly 4's and 5's is an "A"
- Mostly 3's and 4's is a "B"
- Mostly 2's and 3's is a "C"

Handout A4.6,H3, p. 2

661

OR

- To get 100% in writing you have to have 50% or your scores at a 5 with no scores of Ideas and Content, Conventions, or Organization below 4.
- To get 90% in writing you have to have 50% of your scores a 4 or higher with no Ideas and Content, Convention, or Organization below 3, and any other score below a 3 counterbalanced by a score of 4 or higher. Etc.

OR

- To get a C in writing, all writing must be at a 3 or higher. To get an A or a B, students need to choose 5 papers, describe the grade they should get on those papers, and justify the grade using the language of the six-trait model and specific examples from the written work.

  Depending on how the rule finally plays out, Johnny might either get around an A (mostly 4's or 5's) or a B (lots of 4's or 5's, but there are more 4's than 5's) or 90% (there is one 3 in conventions) for the writing part of his grade. Or he might get an A by citing specific examples from the written work and the six-trait rubric that show he really understands what constitutes good writing, and is ready to be a critical reviewer of his own work.

METHOD 2: Total Points. Add the total of possible points students can get on rubric scored papers.

First figure out the number of points possible. To do this, multiply the number of papers being evaluated by the number of traits assessed. Then, multiply that by 5 (or the highest score possible on the rubric). In this case, with 5 papers we would multiply 5 (papers) times 6 (traits) times 5 (highest score on the rubric) and get 150 total points possible.

Then we add a student's scores (Johnny has 109 points) and divide by the total possible—109÷150=0.73. so, Johnny has 73% of possible points, so his writing grade will be 73%. I'll need to weight and combine it with other scores to come up with a single letter grade for the course.

METHOD 3: Total Weighted Points. Add the total of possible points students can get on rubric scored papers weighting those traits deemed most important.

First, figure how many points are possible. You will need to figure out which traits you are weighting. In **Table 2**, assume that we decided to weight Ideas and Content, Organization, and Conventions three times as important as the other three traits. The way to come up with the total points possible is then shown. First, you add all of the scores for each trait (adding the numbers in the column), then multiply the total in each column by its weight. Finally you add up the total in each column to come up with the grand total number of points.

Table 2: Total Possible Weighted Points

Total Possible Points	Ideas and Content	Organization	Word Choice	Sentence Fluency	Voice	Conventions	
Paper 1	5	5	5	5	5	5	
Paper 2	5	5	5	5	5	5	
Paper 3	5	5	5	5	5	5	
Paper 4	5	5	5	5	5	5	
Paper 5	5	5	5	5	5	5	
Totals	25	25	25	25	25	25	
Weights	3	3	1	1	1	3	
Weighted Total	75	75	25	25	25	75	300

Table 3 shows Johnny's scores again and the total using the weighted formula. Johnny has 223 out of 300 weighted points or 74% of total points in writing.

Table 3: Johnny's Writing Scores on Five Papers—With Weighted Totals

Johnny's Scores	Ideas and Content	Organization	Word Choice	Sentence Fluency	Voice	Conventions	
Paper 1	3	2	2	3	1	4	
Paper 2	4	2	3	4	3	4	
Paper 3	4	4	5	5	2	3	
Paper 4	5	4	4	4	2	4	
Paper 5	5	5	5	5	4	4	
Totals	21	17	19	21	12	19	
Weights	3	3	1	1	1	3	
Weighted Total	63	51	19	21	12	57	223

Depending on your focus, you might only want to include the traits you have been working on, weighting others as 0.

METHOD 4: Linear Conversion. Come up with a connection between scores on the rubric and percents, directly. We might find that the wording of a 3 on the six-trait scales comes close to our definition of what a "C" is in district policy, then turn the rubric score into a percent score based on the definition. For example:

1=60%
2=70%
3=80%
4=90%
5=100%

You can then treat the rubric scores the same way you treat other grades in your grade book.

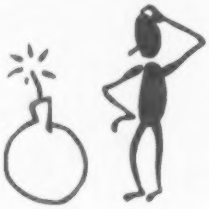
CONCLUSION

One difficulty with the last three approaches is that it makes the method seem more scientific than it is. For example, it is not always clear that the distance between a score of 1 and a score of 2 is the same as the distance between a score of 4 and a score of 5, and linear conversions, or averaging numbers ignores those differences.

Once you, as teacher, arrive at a method for converting rubric scores to a scale that is comparable to other grades, the responsibility is on you to come up with a *defensible* system for weighting the pieces in the grade book to come up with a final grade for students. This part of the teaching process is part of the professional art of teaching. There is no single right way to do it; however, whatever is done needs to reflect evidence of **students' levels of mastery** of the targets of instruction.

Activity 4.7

The Case of the Killer Question



Purposes:

1. To understand the critical link between clearly defined essential learning targets, assessment, grading, and reporting

2. To reflect on and discuss the role of grading in the portrayal of complex learning

Uses:

This is an advanced level activity used in **Chapter 4** to provide an opportunity to participants to apply their understanding of grading and reporting issues. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** Text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); and (b) understanding of how assessment can be integrated with instructions (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*); (c) how assessment fits into standards-based education (**Chapter 1** text and **Activities 1.1**—*Changing Assessment Practices* and **1.12**—*Assessment Principles*; (d) how to match student learning targets to assessment methods (**Chapter 1** text and **Activities 1.2**—*Clear Targets* and **1.7**—*Target-Method Match*; and (e) stage setting activities on grading (**Activity 4.2**—*Putting Grading and Reporting Issues in Perspective* or **4.3**—*Grading Jigsaw* and **Activity 4.5**—*Report Card Formats*).

🗀 Reference Box

The actual site of this case is West Hawaii Explorations Academy. For further information, please contact: Bill Woermer, West Hawaii Explorations Academy, P. O. Box 1749, Kailua-Kona, HI 96745.

Rationale:

This case arises out of an alternative secondary program's dilemma. In the program, self-selected students pursue long-term projects that deepen their knowledge and skills in a very specific area. There are few common content experiences among the program's students. Inquiry, problem solving, persistence, flexibility and other mega-skills are the common strands across student work. Project topics vary significantly from one student or small group to another. Some projects are individual, some are group, all are intended to be viable beyond school boundaries. Student autonomy and self-directed learning are key aims of the program.

As part of a larger high school, however, the program must demonstrate how its students are achieving in specific content areas including language arts, social studies, science, and mathematics for which students receive high school credits. In addition, a significant chunk of the resources that currently support the program come from an external source that emphasizes mathematics, science, and technology. The funding source is another audience expecting information about student learning and the impact of the program on mathematical and scientific literacy.

At the heart of the dilemma is the staff and students' passionate commitment to the independent, integrated, and deep learning embodied in the required projects. They have several key questions to answer: (1) What is the essential learning that the program delivers? (2) How can those outcomes be assessed without losing the complexity of the learning? (3) Do grades make sense in this context? If not, how can learning be reported to the satisfaction of the within-program *and* external audiences?

The use of case studies such as this to stimulate discussion is a useful learning technique for teachers.

Materials Required:

- **Overhead A4.7,O1**—*The Case of the Killer Question Purposes*
- **Handouts A4.7,H1**—*The Case of the Killer Question*, and **A4.7,H2**—*Science and Mathematics Learning Targets for Students*

Time Required:

$1\frac{1}{4}$ to $2\frac{1}{4}$ hours

Facilitator Notes:

1. (2 minutes) Use **Overhead A4.7,O1**—*The Case of the Killer Question Purposes* to describe the goals of the activity.

2. (15-30 minutes) Pass out **Handout A4.7,H1**—*The Case of the Killer Question*. The "killer question" is: *How do we capture and report the learning of our students in ways that others will value without compromising the integration, authenticity, student autonomy, and self-directed learning that is at the heart of our program?*

Participants should read the case and individually reflect on the assessment, grading, and reporting issues embedded in the case, responding to the reflection questions in **Handout A4.7,H1, p. 5**—*Initial Reflections*.

3. (10-20 minutes) Ask participants to establish what's known about the case and invite group members to identify issues and personal connections from their initial reflection. It might be important to determine if there is common understanding of terms like integrated curriculum. Check to see if there are questions that emerged from individual reading and reflection.

4. (30-45 minutes) Use the Discussion Starter questions (**Handout A4.7,H1, p. 6**) to guide and promote discussion of the variety of issues that face this program and its staff. Depending on the relationship of the issues in the case to participants' own local situation, the discussion might move quickly from the program in the case to concerns and practices within a school or district. This is fine. Try to keep the group focused on moving toward actions or solutions that are grounded in quality assessment. One of the expectations of this case is that it will generate and make tangible many often unspoken beliefs about curriculum, assessment, instruction, grading, and reporting.

   Handout A4.7,H2—*Science and Mathematics Learning Targets for Students* can be used to help answer some of the Discussion Starter questions.

5. (10-30 minutes) To bring closure to the discussion (at least for now), have the recorder for each group review key ideas made in the discussion and summarize the group's recommendations for action using **Handout A4.7,H1, p. 7**—*Concluding Thoughts*.

6. (5 minutes) Decide as a group if further study is warranted and set the date for round 2. See the reference box and other activities in *Toolkit98*, **Chapter 4**, for additional ideas and resources.

Reference Box

Kathleen U. Busick and Richard J. Stiggins, *Making Connections: Case Studies for Student-Centered Classroom Assessment, Second Edition*, 1997, Portland, OR: Assessment Training Institute, Inc., (503) 228-3060.

ASCD Year Book: Communicating Student Learning, 1996. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

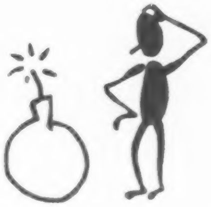
Alfie Kohn, *From Degrading to De-Grading* in Punished by Rewards, 1996, Houghton-Mifflin, (617) 351-5000.

Mike Schmoker, *Results: The Key to Continuous School Improvement*, 1996. Association for Supervision and Curriculum Development (ASCD), (800) 933-2723.

Richard J. Stiggins, *Student Centered Classroom Assessment, Second Edition*, 1997, Merrill.

Ruth Sutton, *Assessment for Learning, 1995*. Author, 29 Nevile Ct., Nevile Rd., Salford, M7 OPS, England UK.

The Case of the Killer Question



Purposes:

1. To understand the critical link between clearly defined essential learnings, assessment, grading, and reporting
2. To reflect on and discuss the role of grading in the portrayal of complex learning

Overhead A4.7,O1

668

The Case of the Killer Question:

How do we capture and report the learning of our students in ways that others will value without compromising the integration, authenticity, student autonomy and self-directed learning that's at the heart of our program?

Context: Images of an alternative secondary program...

It's the end of a typically exhausting day for the staff at an alternative high school program that sits on the site of a national energy laboratory and shares space with lava flows that have hardened into dark waves and swirls. Overhead, airplanes approach the nearby airport, and looking off in the distance it's possible to see some students who have hiked to the ocean at the edge of the lava. The school day is over, but work continues.

The three public school teachers, their partner from the university on the other side of the island, and a staff member who is with the Sea Grant Extension Service are discussing student project proposals which were due today. They are seated on bleachers removed from a junior high school which underwent remodeling; a shaded screening tarp overhead, ripped in places by recent heavy winds, provides some protection from the sun.

Scattered nearby is evidence of projects in varying stages of development. Close by is the frame of a solar car—a reminder of the glory achieved by a group of students in an Australian race commemorated in a Hollywood movie several years ago and the focus of one of the student projects. Off to the left are skeletal frames used in two projects aimed at hydroponics gardening. Two students are hammering on frames to raise the bed on which the newest plants will be placed. Under the tarp are three large aquariums where one student is cultivating sea horses, hoped to be a commercial venture, another is raising shrimp, and a third is raising food for the shrimp. Beyond the tarp toward the ocean is what looks like a pool carved out of the lava. It's not completed yet, but shows signs of readiness for the water that will be piped from the energy lab's deep sea source. To the right are tanks with various reef organisms; these are the focus of Aloha Kai—a program in which the high school students teach elementary youngsters about the reef environment and the living organisms that populate it. Recent research papers turned in by students were part of their preparation to be "expert" teachers on at least one of the reef's inhabitants.

The Academy offers a very different type of learning experience to its students than they would encounter in a traditional high school. The staff are discussing next steps to support and extend the project planning. By this time of year, most students have settled on a major project and have prepared field/lab or literature research that supports their project. A few are still struggling to settle on and define a meaningful project. The teachers are pleased with the bulk of the project proposals but have raised a concern about several students' ability to identify and document the elements of their work that will provide evidence of language arts development, or mathematical literacy. They have worked hard to create an environment which honors and encourages student autonomy and self-directed learning. For some, this means periods of drift and apparent aimlessness. The staff, in response, have developed weekly time management sheets and meet with each student to discuss progress and goals. Daily journals provide insights into the struggles as well as "aha's" of individual students. There are weekly project reviews, individual student binders that house literary critiques, research papers and a computer skills checklist; students also keep portfolios to document their learning in language arts, social studies, science, and mathematics to ensure that they can receive high school credit. Weekly grades are given from work in the weekly folder. Most grades are based on a point system defined in guidelines for various required work.

The student tour guide waves as she heads to a distant site on the energy lab where she's employed in quality control for a company that uses seaweed to provide food supplements. One of the teachers comments that last year at this time, the student was struggling with anger, had trouble zeroing in on a project and at times would walk away if approached by visitors. Her confidence and knowledge of both her own and other students' projects represents a huge leap.

The Dilemma...

The heartfelt pleasure at a week that was productive and challenging gives way among the staff to a discussion of issues that have been surfacing more frequently as they continue to build this integrated program. The high school to which their alternative program is attached wants to know what knowledge and skills in the core content areas are being mastered by students. Some students spend a year and then return to the larger high school and the school wants to know that students are prepared to resume their work in more traditional classes. Student quarterly grades are due to the high school soon. Looking up to talk to a visiting educator, a teacher expresses frustration at having to pick apart small pieces of a very complex whole in order to grade.

At the center of their concern is the issue of core learning. Because students select their own research topics, long term projects and modules for skills development, there appears to be little common content.

670

Handout A4.7,H1, p. 2

Teachers Voices...

> *How do we talk about the mathematics that's needed to construct a hydroponics structure, determine spacing and nutrient needs for the plants (in one instance ornamental flowers, in another vegetables) grown within this area and try to compare it to the mathematics involved in determining what temperature the water in our reef tank needs to be in order to sustain the various organisms; or to estimate and then test the amount of brine shrimp needed to support other aquarium life? We have students who are beginning from scratch, building the environment for a lava bed tide pool; figuring out how to safely transport reef organisms from the school site to the energy lab where elementary students are taken on tour. The knowledge needed for each is so different. How do all these very different things add up to a single grade for each student? And what will that grade tell anyone about the actual things our students know and can do? How can a grade communicate to others the transformation that has taken place for students who are now selecting projects, pursuing them, constructing the environment they need, negotiating with mentors for additional information and support–that's just the beginning of a long list of what's happening here.*
>
> *We know right now how each student is doing. It's not a mystery. We see them working every day and have lots of products before our eyes. But how do we tease out the specific learning in each student's work? And after that, do grades make any sense? We see differences in attitude, differences in the challenges they set and accept for themselves, differences in the thinking and planning that go into their work. It's like we're trying to reduce something very real and multifaceted into nice little boxes with single labels.*

Searching for Solutions...

Working together, the teachers developed a list of questions that they'll be tackling together in preparation for a discussion with high school administrators and a visit from an assessment specialist whose role is to provide them with assistance as they rethink and respond to the challenges of grading and reporting their students' learning:

- What are the **common** elements of our students' learning? Are we assessing them well?
- **Why** should we try to reduce complex learning to a single symbol?

Handout A4.7,H1, p. 3

- Is it essential that there be **common subject content** that we assess or does that choice subvert our mission?

- How can we (and our students) **honestly portray** and **anchor** our reporting of their learning from unique, integrated, and often individual experiences?

- How well does our current grading **inform our own decisions**?

- How well does it **inform our non-program users** (the high school, other teachers, colleges, etc.) when it is based on experiences that are not common to all of our students?

- **Do grades make real sense** for us and our program? What do they convey that we can't show some other way?

- **How are our students best served?**

Case discussion tasks...

There are two steps ahead:

1. As the assessment specialist called in to provide assistance and help the staff honor their vision for the program while providing for the information needs of others, begin by writing your initial reflections. Use this time to clarify the key issues and challenges you see in the case, seek personal connections between the case and your own experiences, and consider the possible concerns that will need to be dealt with in the study group discussion in order to reach agreement on the final question.

2. Meet as a study group to discuss the case and propose solutions to the dilemma facing the staff.

672

Handout A4.7,H1, p. 4

Initial Reflections...

Freewrite your initial responses to this case or use the questions below to guide your reflection and preparation for the group discussion coming up.

- What challenges does this case present?
- What kinds of learning targets is this program aiming for?
- What's your initial reaction to the concerns of the high school?
- Do you have some experiences related to this case that can contribute to the discussion?

Discussion Starters...

- What do we know about this case? What are the facts?
- How do we deal with things that are highly valued but whose definition is vague and perhaps different for each student?
- How might the staff more clearly define the learning targets?
- What assessment implications do you see as this group continues its work?
- What actions should this staff take with respect to assessment? What do you recommend?
- When it comes to grades, what do you say to these teachers? What actions do you urge them to take? What actions do you take, given your role?

Concluding Thoughts...

Summarize your group's discussion by answering the question that began this case:

> How do we capture and report the learning of our students in ways that others will value without compromising the integration, authenticity, student autonomy and self-directed learning that's at the heart of our program?

Handout A4.7,H1, p. 7

Science and Mathematics Learning Targets for Students

After reviewing several standards documents including the *Pacific Standards for Excellence in Science and Mathematics*, the *Hawaii Content and Performance Standards*, and the *New Standards Consultation Draft Performance Standards*, the staff identified a preliminary set of essential learnings that their program addresses.

I. Science

A. Physical sciences concepts

1. The student understands chemical reactions, including concentration, pressure, temperature, and catalysts

B. Life sciences concepts

1. The student understands interdependence of organisms, especially flow of energy, cooperation and competition, and environmental constraints
2. The student understands biological evolution, in particular, natural selection; and adaptation, including species, variation, and extinction

C. Earth and space sciences concepts

1. The student understands natural resource management

D. Science connections and applications

1. The student understands big ideas and unifying concepts; for example, order and organization, models, systems, evolution and equilibrium, form and function, cause and effect, and constancy and change
2. The student understands technology, including cost/benefit, constraints, feedback, and risk
3. The student understands the designed world, including agriculture and industry
4. The student understands historical and contemporary impact of science

E. Scientific thinking—The student uses scientific reasoning strategies, scientific knowledge, and common sense to formulate questions about, understand, and explain a wide range of phenomena; that is:

1. The student identifies problems or design opportunities; proposes designs and chooses among alternatives; implies a solution and evaluates its consequences
2. The student works individually and in teams to collect and share information and ideas

F. **Scientific tools and technologies**—The student uses tools and technologies to collect and analyze data; that is:

1. The student records and stores data in a variety of formats, including databases, audio tapes, and videotapes
2. The student acquires information from print, electronic and visual sources, including the Internet

G. **Scientific communication**—The student communicates clearly and effectively about the natural world; that is:

1. The student represents data and results in multiple ways; for example, numbers and statistics; drawings, diagrams, and pictures; sentences; charts and tables; models; and uses the most effective way to make the point
2. Summarizes varied sources of evidence including his or her own data and published reports

H. **Scientific investigation**

1. The student completes projects drawn from the following types of investigation, including at least one full investigation each year and, over the course of high school, investigations representing all four types:
    a. controlled experiment
    b. fieldwork
    c. design
    d. secondary research, i.e., use of others' data

2. A single project may draw on more than one type of investigation. A full investigation includes:
    a. questions that can be studied using the resources available
    b. procedures that are safe, humane, and ethical; respect privacy and property rights
    c. data that have been collected and recorded in ways that others can verify, and analyzed using skills expected at this grade level
    d. data and results that have been represented in ways that fit the context
    e. recommendations, decisions, and conclusions based on evidence
    f. acknowledgment of references and contributions of others
    g. results that are communicated appropriately to audiences
    h. reflection and defense of conclusions and recommendations from other sources and peer review

Handout A4.7,H2, p. 2

II. Mathematics

A. Number and operation concepts

1. The student uses the properties of addition, subtraction, multiplication, division, exponentiation, and root-extraction in forming and working with algebraic expressions
2. The student uses dimensionless numbers, such as proportions, percents, and multiplicative factors; and numbers with specific units of measure, including length, time and rate units

B. Geometry and measurement concepts

1. The student knows, uses, and derives formulas for areas, surface area, and volume of many types of figures
2. The student works with geometric measures of length, area, surface area, volume, and angle; and non-geometric measures of weight, monetary value, and time

C. Statistics and probability concepts

1. The student formulates hypotheses to answer a question and uses data to test hypotheses

D. Problem solving and mathematical reasoning—The student solves problems that make significant demands in one or more of these aspects of the solution process: problem formulation, problem implementation, and problem conclusion

1. Problem formulation
    a. The student participates in the formulation of problems; in particular, given the basic statement of a problem situation:
        (1) the student fills out the formulation of a definite problem that is to be solved
2. Problem Implementation
    a. The student makes the basic choices involved in planning and carrying out a solution; in particular
        (1) the student chooses and employs effective problem-solving strategies in dealing with non-routine and multi-step problems
        (2) the student applies mathematical concepts to new situations within mathematics and uses mathematics to model real world situations involving basic applications of mathematics in the physical sciences, social sciences, and business

Handout A4.7,H2, p. 3

3. Problem conclusion
    a. The student provides closure to the solution process through summary statements and general conclusions; in particular
        (1) the student concludes a solution process with a useful summary of results

E. Mathematical skills and tools

1. The student makes reasonable estimates in appropriate units of quantities met in applications
2. The student makes and uses rough sketches, schematic diagrams, or precise scale diagrams to enhance a solution
3. The student uses technology to create graphs or spreadsheets that contribute to the understanding of a problem

F. Mathematical communication

1. The student uses mathematical representations with appropriate accuracy, including numerical tables, formulas, functions, algebraic equations, charts, graphs, and diagrams

G. Putting mathematics to work

1. The student conducts at least one large scale investigation or project each year drawn from the following kinds and, over the course of high school, investigations or projects drawn from at least three of the kinds. A single investigation or project may draw on more than one kind.

    a. Data study in which the student:
        (1) carries out a study of data relevant to current civic, economic, scientific, health, or social issues
        (2) prepares a report that explains the purpose of the project, the organizational plank, and conclusions, and uses an appropriate balance of different ways of presenting information

    b. Design of a physical structure, in which the student:
        (1) uses general mathematical ideas and techniques in discussing specifications for building the structure
        (2) prepares a report that explains the purpose of the project, the organizational plank, and conclusions, and uses an appropriate balance of different ways of presenting information

Handout A4.7,H2, p. 4

c. Management and planning analysis, in which the student:
   (1) carries out a study of a business or public policy situation involving issues such as optimization, cost-benefit projects, and risk
   (2) uses decision rules and strategies both to analyze options and balance tradeoffs; and brings in mathematical ideas that serve to generalize the analysis across different conditions
   (3) prepares a report that explains the purpose of the project, the organizational plank, and conclusions, and uses an appropriate balance of different ways of presenting information

Handout A4.7,H2, p. 5

Activity 4.8
Guidelines for Grading

Goals:

1. To think through purposes for grading (and other mechanisms for reporting student achievement and progress
2. To think through guidelines for grading and other reporting that will assist us accomplish our purposes

Uses:

This is an advanced level activity. Prerequisites include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); (b) understanding of how assessment can be integrated with instruction (**Chapter 2** text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*); and (c) a stage-setting activity in which grading and reporting issues are discussed, such as **Activity 4.3**—*Grading Jigsaw* or **Activity 4.2**—*Putting Grading and Reporting Questions in Perspective*.

Rationale:

In most schools, at least for the foreseeable future, students will receive report card grades. If grading must be done, we all must try to provide *better* grades, that is, grades which:

- Encourage learning and support student success
- Are not punitive
- Are meaningful and provide useful information

This activity presents eight guidelines for grading, taken from current research and thinking about grading. Participants have the opportunity to discuss the advisability of each guideline and how it might or might not contribute to the three criteria listed above.

🕮 Reference Box:

Greater detail on these guidelines can be found in an article by Ken O'Connor in the May 1995 NASSP Bulletin, and in a forthcoming book from IRI/Skylight called *Good Grades: Grading Practices Which Encourage Learning.* (800) 348-4474, e-mail: info@iriskylight.com

681

Materials:

- Overhead projector, screen, overhead pens, blank transparencies
- **Overheads A4.8,O1**—*Guidelines for Grading Purposes,* **A4.8,O2**—*Should Grades ….* and **A4.8,O3**—*Discussion Questions*
- **Handout A4.8,H1**—*Grading Guidelines*

Time Required:

90 minutes to two hours

Facilitator's Notes:

1. (5 minutes) Use **Overhead A4.8,O1**—*Guidelines for Grading Purposes* to introduce the purpose for the activity.
2. (10-15 minutes) Ask participants to discuss the proposition that grades should (see **Overhead A4.8,O2**—*Should Grades…*):
   - Encourage learning and support student success
   - Not be punitive
   - Be meaningful and provide useful information

   The rest of the activity depends on agreement on these purposes for grading. If participants cannot agree on the purposes for and roles of grading in the educational process, then it will become more difficult to develop guidelines for grading.
3. (10-15 minutes) Break participants into small groups. Assign each group one or more guideline(s) (**Handout A4.8,H1**—*Grading Guidelines*) to discuss (see **Overhead A4.8,O3**—*Discussion Questions*):
   - What do you like about this guideline?
   - What makes you uncomfortable about this guideline?
   - What points of uncertainty do you have about this guideline?
   - To what extent will this guideline assist in accomplishing our purposes for grading?
4. (45 minutes) Share ideas and thoughts in the whole group.
5. (15-30 minutes) Ask participants to develop their own list of guidelines for grading.

Grading Guidelines

Purposes:

1. To think through purposes for grading
2. To think through guidelines for grading that will assist us accomplish our purposes

Overhead A4.8,O1

Should Grades...

… encourage learning and support student success?

… not be punitive?

… be meaningful and provide useful information?

684

Overhead A4.8,O2

Discussion Questions

- What do you like about this guideline?
- What makes you uncomfortable about this guideline?
- What points of uncertainty do you have about this guideline?
- To what extent will this guideline assist in accomplishing our purposes for grading?

Overhead A4.8,O3

Guidelines for Grading–
to support learning and encourage student success

1. **Limit the valued attributes included in grades to individual achievement.**
   a. Individual achievement is the only basis for grades (not effort, behavior or tardiness).
   b. Effort, participation, attitude and other behaviors should be reported separately.

2. **Sample student performance—don't include all work in the summary evaluation (mark or grade).**
   a. Provide feedback on all performance.
   b. Include only summative performances in grades, not homework or practice work..

3. **Grade in pencil.**
   a. Use the most recent information.
   b. Provide several assessment opportunities.

4. **Relate grading and reporting procedures to the intended learning (outcomes/targets).**

5. **Use care "crunching" numbers.**
   a. The average doesn't have to be the mean; consider using medians.
   b. Weight information properly to achieve intent in final grades.

6. **Use criterion-referenced (i.e., preset) standards to distribute grades (and marks).**

7. **Properly record evidence from quality assessment instruments.**

8. **Discuss assessment and evaluation practices, including grading with students at the beginning of instruction.**

Adapted from Ken O'Connor, Scarborough Board of Education, Ontario, Canada

Handout A4.8,H1

686

Activity 4.9
Grading Scenarios

Purpose:

To use real grading situations to raise and discuss grading issues and solutions

Uses:

This activity is at an advanced level—it can be thought of as the application of the knowledge and perspectives generated through previous activities. Prerequisites might include: (a) knowledge of how student assessment fits into standards-based education (**Chapter 1** text and/or **Activities 1.1**—*Changing Assessment Practices*, or **1.12**—*Assessment Principles*); (b) understanding of how assessment can be integrated with instruction (**Chapter 2** Text and/or **Activities 2.1**—*Sorting Student Work* and **2.3**—*Ms. Toliver's Mathematics Class*); (c) how assessment fits into standards-based education (**Chapter 1** text and **Activities 1.1**—*Changing Assessment Practices* and **1.12**—*Assessment Principles*; (d) how to match student learning targets to assessment methods (**Chapter 1** text and **Activities 1.2**—*Clear Targets* and **1.7**—*Target-Method Match*; and (e) stage setting activities on grading (**Activity 4.2**—*Putting Grading and Reporting Issues in Perspective* or **4.3**—*Grading Jigsaw* and **Activity 4.5**—*Report Card Formats*).

Rationale:

This activity is about the validity of grades, specifically the kind of information that should go in a grade, how it should be weighted, and how related matters might be handled. Participants are presented with seven real-life grading situations; they have to decide what to do and provide a rationale for their chosen course of action. Three scenarios are about effort relative to ability, two are about missing work, and two are about improvement.

Part of the discussion will relate to a difference of opinion about what constitutes "achievement." For many teachers, turning work in on time is a sort of "achievement" that they deliberately try to teach. For others, it's not. Part of the discussion, especially about the "improvement" scenarios, will deal with the difference between criterion-referenced and self-referenced grading. These scenarios will also bring up issues about grading that relate to Alfie Kohn's Level III concerns—why do we grade? Is grading the best way to accomplish motivation, feedback, and reward? (See Article #1 in **Appendix C**—*Articles)*

The scenarios have been used in research with teachers. The percentage of teachers in the research study choosing each solution is provided in **Overhead A4.9,O3**—*Research Results*.

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens
- **Overheads A4.9,O1**—*Grading Scenarios Purposes*, **A4.9,O2**—*Grading Scenarios Instructions*, and **A4.9,O3**—*Research Results*
- **Handouts A4.9,H1**—*Grading Scenarios*, and **A4.9,H2**—***Scenario** Discussion Sheets* (one per person)

Time Required:

90 minutes

Facilitator's Notes:

1. (5 minutes) Use **Overhead A4.9,O1**—*Grading Scenarios Purposes* to review the purposes for this activity.

2. (35 minutes) Either individually or in small groups, participants should decide which choice they think is the best for each grading scenario (**Handout A4.9,H1**—*Grading Scenarios*) and decide WHY they think that is the best choice. They can use **Handout A4.9,H2**—***Scenario** Discussion Sheet* to guide their thinking. (Instructions are also on **Overhead A4.9,O2**—*Grading Scenario Instructions*.)

3. (35 minutes) Facilitate a discussion of each scenario. One procedure would be to assign one group to discuss each scenario and then ask other groups to elaborate if they have anything to add. (The responses of one group to Scenarios 2, 3, 4, and 5 are included below for reference. **Overhead A4.9,O3**—*Research Results* shows the responses of a group of practicing teachers.)

   Challenges to dubious grading practices are best if they come from peers. If they have to come from you (the facilitator) because no one else brings it up, phrase them as questions (e.g., "What kind of information are you using if you grade that way?" or "What would your unit learning targets have said were the things they should be evaluated on?")

4. (15 minutes) Journaling activity. Ask participants to individually write down the five most important things they would like to remember from this discussion, and why they want to remember them.

Sample Responses From A Previous Group:

Scenario Discussion Sheet–Scenario 2

Issues:

1. Averaging a "0" in with other scores unfairly lowers an average.
2. What is role of homework? If the role is practice, then why include it in a grade?
3. What is the grade supposed to reflect—amount of work done, effort, or knowledge level?
4. The grade is only as good as the assessments on which they are based. Which is a better measure of achievement in this case—homework or tests/quizzes?

What might be missing in the teacher's grading scheme for this situation to have arisen in the first place?

1. How *good* are the tests and quizzes? What is a "working out exercise?" Do the tests and quizzes really reflect what students know and can do? If so, then the tests and quizzes provide the best basis for a grade.
2. What are the criteria for grading? Are these sound?

What advice would you give this teacher?

There really wasn't enough information to decide what to do. The final decision depends on the student—Is this a special needs student? What is the reason for the lack of homework? Was it done and just not turned in?

What percentage of practicing teachers do you think made each choice?

a. 65%
b. 10%
c. 10%
d. 15%

Scenario Discussion Sheet–Scenario 3

Issues:

1. What is the goal of grading in this class—to reward effort or provide feedback on achievement?

2. Grading criteria should be fairly applied across students.

3. What does a grade mean if, for every student, ability is balanced against actual performance? Could smart students ever get "A's"?

What might be missing in the teacher's grading scheme for this situation to have arisen in the first place?

1. What is the quality of instruction? Are the student's learning needs addressed? Is the course engaging?

2. Is the teacher clear about the learning targets he has for students and how all the assessment that is collected relates to these targets?

What advice would you give this teacher?

Give the student the grade based on quality of work, not effort or achievement in relationship to ability. If anything else is to be done, it should be set up in advance and everyone should know the criteria.

What percentage of practicing teachers do you think made each choice?

a. 90%
b. 1%
c. 0%
d. 9%

Scenario Discussion Sheet–Scenario 4

Issues:

1. Why would a student ever be put in this position?

2. The real decision is whether the student has the competencies, not the number of points. If the student has the competencies, they should pass.

3. What is the grade supposed to reflect—amount of work done, effort, or knowledge level?

4. The grade is only as good as the assessments on which they are based. Do the tests really reflect competency?

5. There could be more than one standard of comparison—criterion-referenced, norm-referenced and self-referenced. Everyone in the class needs to know the rules from the outset.

What might be missing in the teacher's grading scheme for this situation to have arisen in the first place?

1. How *good* are the tests and quizzes; do they really reflect what students know and can do?
2. Perhaps assign more than one grade—criterion-referenced, norm-referenced, and self-referenced.

What advice would you give this teacher?

The teacher needs to go back and determine the strengths and weaknesses of the student and the extent to which the tests really reflect achievement.

What percentage of practicing teachers do you think made each choice?

a. 70%
b. 10%
c. 20%

Scenario Discussion Sheet–Scenario 5

Issues:

1. Is instruction an issue here? Could the student have been more motivated if instruction was different? How do you define effort? Is lack of effort due to the student or teacher?
2. What is the grade supposed to reflect—amount of work done, effort, or knowledge level?
3. The grade is only as good as the assessments on which they are based. Do the assessments really measure what students know and can do?

What might be missing in the teacher's grading scheme for this situation to have arisen in the first place?

Effort should count, but it should be handled separately from achievement.

What advice would you give this teacher? Choice B

What percentage of practicing teachers do you think made each choice?

a. 0%
b. 100%
c. 0%
d. 0%

Grading Scenarios

Purpose:

Use real grading situations to raise and discuss grading issues and solutions.

Overhead A4.9,O1

Grading Scenarios
Instructions

Discuss the scenario in your group. Be ready to describe:

- The scenario
- The issues raised by the scenario
- What might be missing in the teacher's grading scheme to have given rise to this problem to begin with
- The advice you'd give the teacher
- The percentage of teachers you believe would choose each response

Overhead A4.9,O2

693

Grading Scenarios—Actual Teacher Responses

Scenario 1 a. 1% b: 51% c. 48%	Scenario 4 a. 94% b. 6%
Scenario 2 a. 86% b. 13% c. 1%	Scenario 5 a. 5% b. 95% c. 0%
Scenario 3 a. 81% b. 16% c. 4%	Scenario 7 a. 73% c. 27%

🗀 **Reference Box:**

Adapted from: Susan M. Brookhart, 1993. "Teacher's Grading Practices: Meaning and Values." *Journal of Educational Measurement 30*, pp. 123-142.

Note: The questions in this activity have been revised from those used in the research. For example, teachers in the study were not given a "something else" choice. Question 6 was revised to such an extent that responses from the study would not be meaningful. Question 7 originally had only two responses—we've added "b" to our version.

Overhead A4.9,O3

Grading Scenarios

Each of the following is a scenario that describes a decision a teacher has to make about assigning a grade. The teacher is unsure of what to do and has asked you for some advice. What advice would you give? What issues do each choice raise? What questions need to be considered before making a decision? What might be missing in a teacher's grading process for the situation to arise at all?

Scenario 1

In Mr. Nguyen's seventh grade social studies class, report card grades were based on quizzes, tests, and an out-of-class project which counted as 25% of the grade. Terry obtained an A average on his quizzes and tests but has not turned in the project despite frequent reminders. In this situation, Mr. Nguyen should:

a. Exclude the missing project and give Terry an A.
b. Assign Terry a 0 for the project and a F on his report card because his average would be 68%.
c. Assign Terry a C, counting off some for not turning in the project.
d. Something else…

Scenario 2

Mr. Marlowe's 9th grade class has students with varying ability levels. During this grading period, the students' grades are based on quizzes, tests, and homework assignments which involve working out exercises. Kelly has not turned in any homework assignments despite Mr. Marlowe's frequent reminders. Her grades on the quizzes have ranged from 65% to 75% (C to D), and she received a D on each of the tests. In this situation, Mr. Marlowe should:

a. Assign Kelly a 0 for the homework assignments and include this in the grade, thus giving her an F for the grading period.
b. Ignore the missing homework assignments and give Kelly a D.
c. Ignore the missing homework and give Kelly a C.
d. Something else…

Handout A4.9,H1, p. 1

Scenario 3

Mr. Paderewski is the teacher in a 6th grade heterogeneously grouped class. Chris, one of his students, has high academic ability, as shown by her previous work, test results, reports of other teachers, and his own observation. As he looks over her work for the grading period, he realizes that the quality of her work is above average for the class, but the work does not represent the best that she could do. The effort she has shown has been minimal, but, because of her high ability, her work has been reasonably good. In this situation, Mr. Paderewski should:

a Grade Chris on the quality of her work in comparison to preset standards, without being concerned about the amount of work that she could have done.
b. Lower Chris's grade because she did not make a serious effort in this class; she could have done better.
c. Assign Chris a higher grade to encourage her to work harder.
d. Something else…

Scenario 4

Ms. Quantum has a heterogeneously grouped 7th grade science class. Barbara is one of her lower performing students, as measured by previous performance and the observations of her former science teachers. Throughout this grading period Barbara has worked very hard. She has turned in her assignments on time and has often gone to Ms. Quantum for extra help before tests. Her average for this grading period is 2 points below what is needed for a D on Ms. Quantum's grading scale. In this situation, Ms. Quantum should:

a. Give Barbara a D for the effort she has shown.
b. Grade Barbara according to the work she has done and assign her an F.
c. Something else…

Scenario 5

Ms. Yohansen has a heterogeneously grouped 5th grade class. Sandy, a student in her class, appears to have average ability to do the required work. In evaluating her work for this grading period, Ms. Yohansen observed that Sandy did not do the work she is capable of—she could have done better. Sandy, however, managed to meet what is required for her to get a C. In this situation, Ms. Yohansen should:

a. Assign Sandy a lower grade because she could have put more effort into the class and could have done better.
b. Assign Sandy a grade based on the quality of her work, without taking into account the amount of work that she could have possibly put in.
c. Assign Sandy a higher grade to encourage her to try harder.
d. Something else…

Scenario 6

Ms. Exponent is teaching high school algebra. In her class of general and academic track students, she gives two tests each grading period. David got an F on the first test and a low B on the second. In this situation, Ms. Exponent should:

a. Assign David an overall grade of D based on the average of his performance on the two exams.
b. Assign David an overall grade of C because he showed improvement in his performance.
c. Assign David an overall grade of B because that was his level of performance at the end of the term.
d. Something else…

Scenario 7

Mr. Phylum is a high school introductory biology teacher. In this class he gives two exams each term. In calculating Bernie's grade for this term, he notices that on the first exam, Bernie obtained a score equivalent to a A and on the second exam, a low C. In this situation, Mr. Phylum should:

a. Assign Bernie an overall grade of A, giving him the benefit of the doubt.
b. Assign Bernie an overall grade of B which is the average of his scores on the two exams.
c. Assign Bernie an overall grade of C, noting that there was decline in his performance.
d. Something else…

Handout A4.9,H1, p. 3

🗀 Reference Box:

Adapted from: Susan M. Brookhart, 1993. "Teacher's Grading Practices: Meaning and Values." ***Journal of Educational Measurement 30,*** **pp. 123-142.**

Scenario Discussion Sheet

Issues:

What might be missing in the teacher's grading scheme for this situation to have arisen in the first place?

What advice would you give this teacher?

What percentage of practicing teachers do you think made each choice?

Handout 4.9,H2